Correlation
Power analysis
Analysis of variance (ANOVA)
Multiple hypothesis testing

Biostatistics Course 2024
Lecture 4
Thursday, 11 July 2024
10:00pm - 12:00pm

Correlation

Example: lipids and insulin sensitivity

sensitivity	fatty_acid
250	17.9
220	18.3
145	18.3
115	18.4
230	18.4
200	20.2
330	20.3
400	21.8
370	21.9
260	22.1
270	23.1
530	24.2
375	24

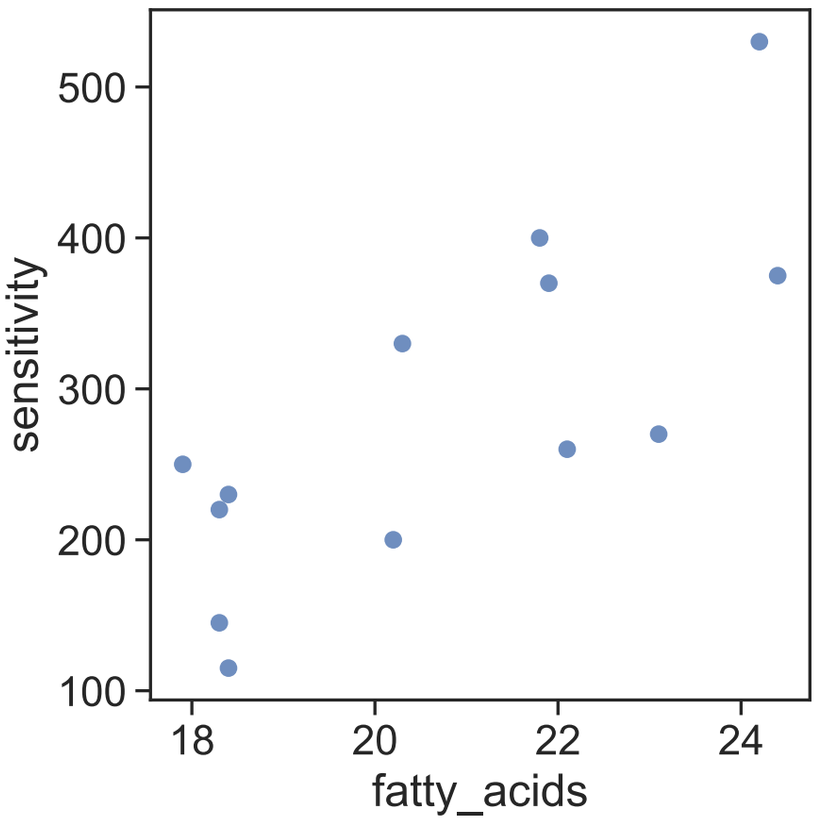
Borkman et al. (1993) wanted to understand why insulin sensitivity varies so much among individuals. They hypothesized that the lipid composition of the cell membranes of skeletal muscle affects the sensitivity of the muscle for insulin.

They determined the insulin sensitivity of $N = 13$ healthy men by infusing insulin at a standard rate (adjusting for size differences) and quantifying how much glucose they needed to infuse to maintain a constant a blood glucose level…

They also took a small muscle biopsy from each subject and measured its fatty acid composition. We'll focus on the fraction of of polyunsaturated fatty acids that have between 20 and 22 carbon atoms ("fatty_acid").

Correlation is used to describe relationships between real-numbered variables

- a measure of relatedness of two variables, X and Y
- independent of measurement units
- ranges between -1 and 1



summary statistics

	pearson
N	13
r	0.77
95% CI	[0.38, 0.93]
r^2	0.593
P-val	0.00207701

Covariance and correlation are estimated from data in the familiar manner

The formula for variance is

$$\widehat{\text{var}}(x) = \sigma_x^2 = \frac{1}{N-1}\sum_i (x_i - \hat{\mu}_x)^2$$

Covariance is estimated in a manner similar to variance

$$\widehat{\text{cov}}(x, y) = \frac{1}{N-1}\sum_i (x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y)$$

The corresponding “correlation coefficient” is

$$r = \frac{\widehat{\text{cov}}(x, y)}{\hat{\sigma}_x \, \hat{\sigma}_y}$$

https://en.wikipedia.org/wiki/Pearson_correlation_coefficient

This is what the correlation coefficient looks like

Pearson's r ranges from -1 to 1.

$r = 0$ implies independence or no relationship, i.e. $p(x, y) = p(x) \cdot p(y)$.

$r = \pm 1$ when the two variables share a deterministic linear relationship.

r close to 1 implies nearly perfect positive dependence

r close to -1 implies nearly perfect negative dependence

Adding a constant to all x or all y, or a multiplicative rescaling of all x or all y, do not change r.

https://en.wikipedia.org/wiki/Pearson_correlation_coefficient

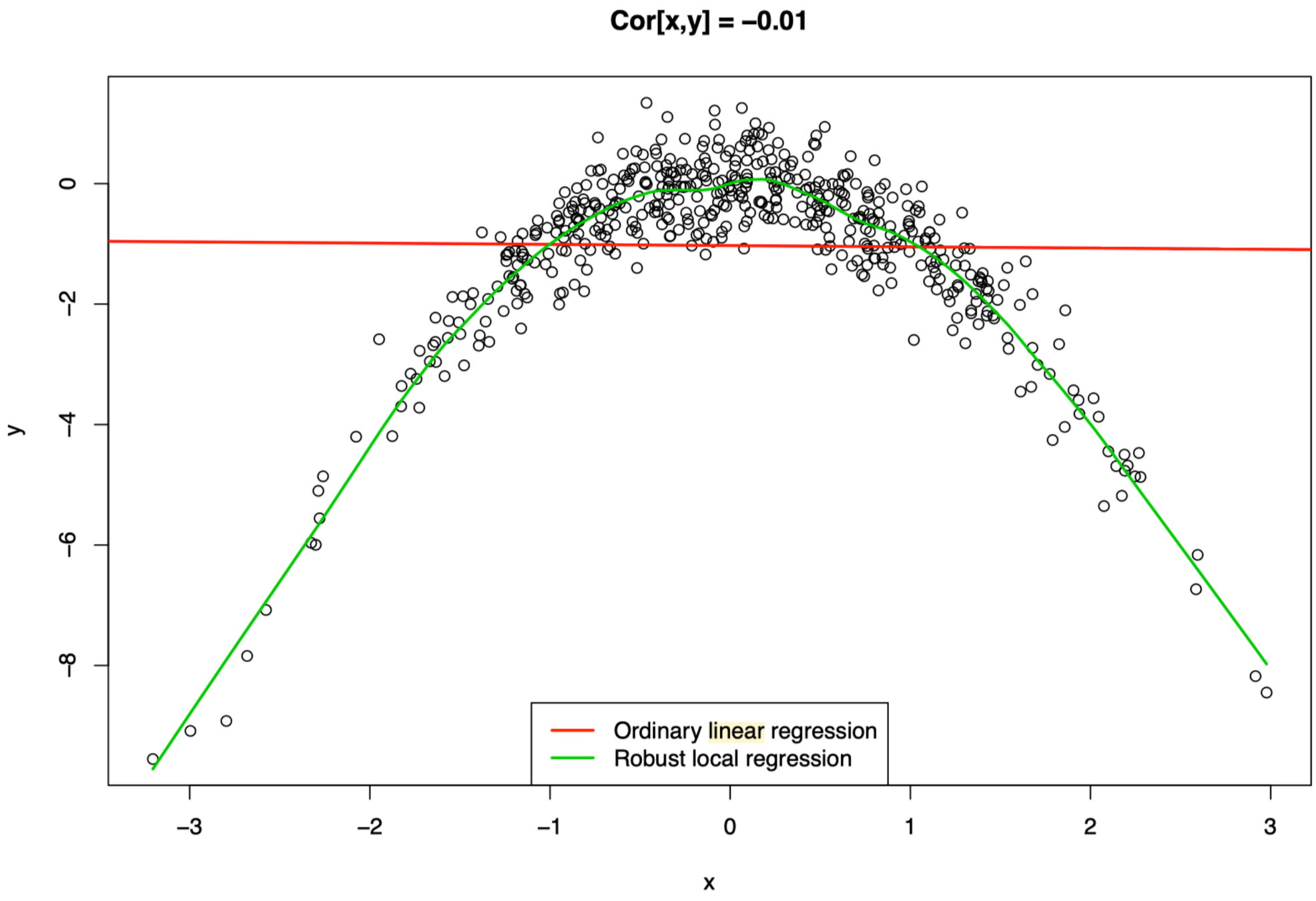
This is what the correlation coefficient looks like

In the deterministic case, r is unaffected by the magnitude of the slope relating two variables, while the sign of r is equal to the sign of the slope.

Sometimes $r = 0$ when two variables have a non-linear relationship. Note that the correlation coefficient only captures **linear relationships** between two variables.

https://en.wikipedia.org/wiki/Pearson_correlation_coefficient

Example: Quadratic Association



The coefficient of determination another name for r^2

The coefficient of determination is simply r^2, which is also often written as R^2.

r^2 is always between 0 and 1 (inclusive)

Remember that $r^2 \leq |r|$, so beware of people reporting r instead of r^2 to make a correlation seem stronger.

r^2 is commonly interpreted as the fraction of variance in y explained by x (or the other way around).

Hypothesis testing

Null hypothesis is “no correlation between the variables”

$$H_0 : \rho = 0$$

Alternative hypothesis is “there is a relationship between the variables”

$$H_a : \rho \neq 0 \qquad \text{(two-sided), or}$$

$$H_a : \rho < 0 \qquad \text{(one-sided less, or)}$$

$$H_a : \rho > 0 \qquad \text{(one-sided greater)}$$

Test statistic is t-statistic that has a t_{n-2} under the null hypothesis

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Hypothesis testing

Null hypothesis is “no correlation between the variables”

$$H_0 : \rho = 0$$

Alternative hypothesis is “there is a relationship between the variables”

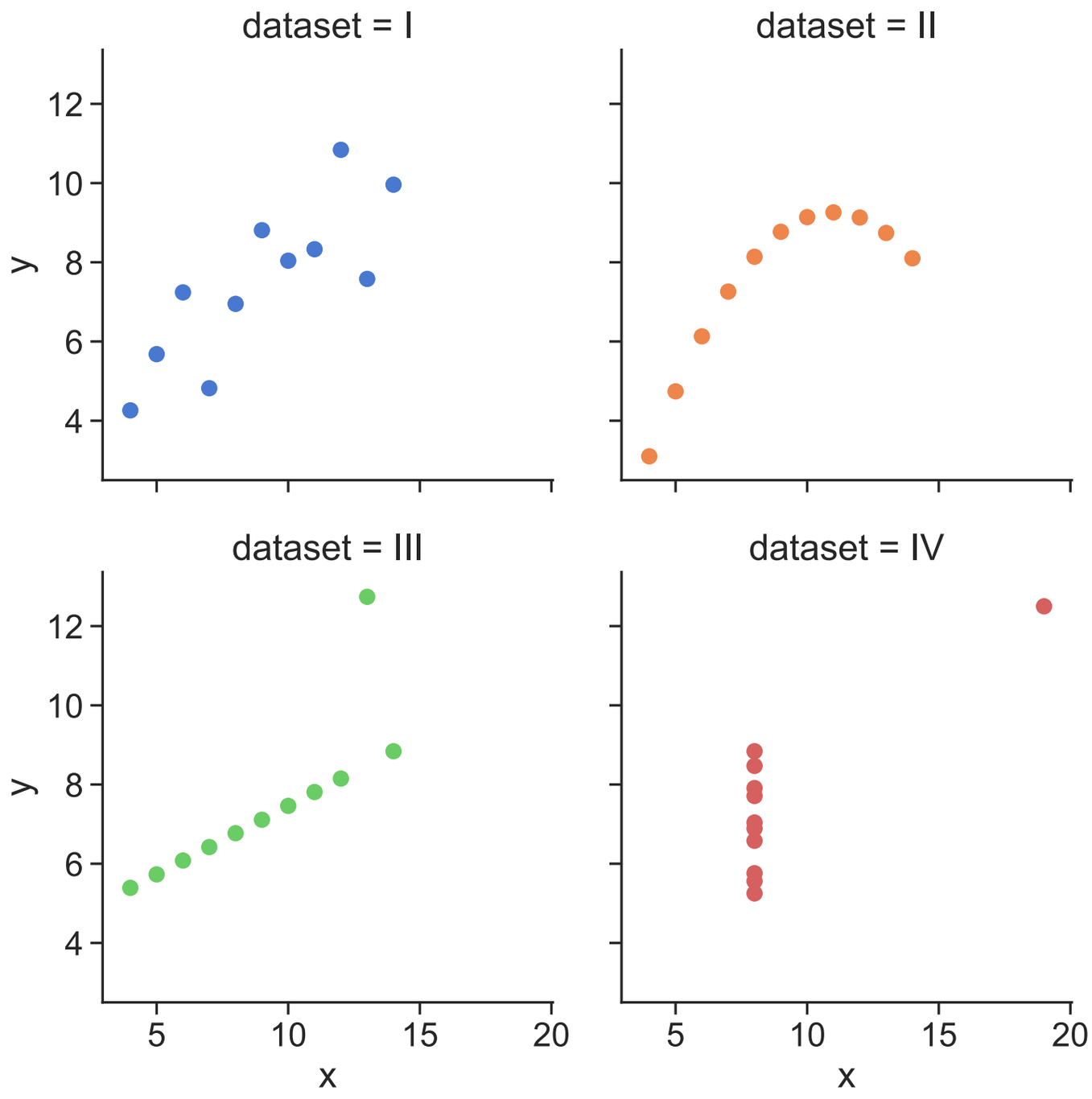
$H_a : \rho \neq 0$ (two-sided), or

$H_a : \rho < 0$ (one-sided less, or)

$H_a : \rho > 0$ (one-sided greater)

Lots of different-looking datasets will have the same value for r.

"Anscombe's quartet": $r = 0.816$ for all 4 datasets



Anscombe, F. J. (1973). "Graphs in Statistical Analysis". American Statistician. 27 (1): 17–21.

Assumptions underlying correlation

Interpreting the correlation coefficient r, and especially the associated P-value, requires multiple assumptions:

- Each data point (x, y) is independently sampled from a 2D Gaussian distribution.
- In particular, x and y each follow a 1D Gaussian distribution
- All covariation between x and y is **linear**, with perfect concordance disrupted only by Gaussian noise.

There are usually many explanations for why two variables might correlate

Possible reasons for a correlation between lipid levels and insulin sensitivity:

- The lipid content of membranes affects insulin sensitivity
- The insulin sensitivity affects membrane lipid content
- Both insulin sensitivity and lipid content are under the control of some third factor, such as a hormone.
- Lipid content, insulin sensitivity, and other factors are all part of a complex molecular/biochemical/physiological network, perhaps with positive and/or negative feedback components. The correlation observed is just a peak at a much more complex set of interdependent relationships.
- Membrane lipid content and insulin sensitivity don't actually correlate at all; the result is just a coincidence.

Correlation is NOT causation!!!

Welcome to GraphPad Prism

GraphPad Prism
Version 8.2.1 (279)

NEW TABLE & GRAPH

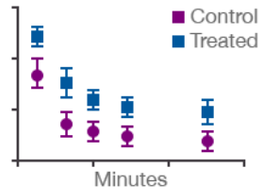
- XY
- Colum
- Group
- Contingency
- Survival
- Parts of Whole
- Multiple variables
- Nested

EXISTING FILE

- Open a File
- LabArchives
- Clone a Graph
- Graph Portfolio

XY tables: Each point is defined by an X and Y coordinate

		X	A			B		
		Minutes	Control			Treated		
		X	A:Y1	A:Y2	A:Y3	B:Y1	B:Y2	B:Y3
1	Title							
2	Title							
3	Title							



? Learn more

Data table:

- (•) Enter or import data into a new table
- ( ) Start with sample data to follow a tutorial

Options:

X:
- (•) Numbers
- ( ) Numbers with error values to plot horizontal error bars
- ( ) Dates
- ( ) Elapsed times

Y:
- ( ) Enter and plot a single Y value for each point
- ( ) Enter 3 replicate values in side-by-side subcolumns
- ( ) Enter and plot error values already calculated elsewhere

Enter: Mean, SD, N

Prism Tips

Cancel

Create

correlation.pzfx

Search

- Data Tables
  - **Data 1**
  - *New Data Table...*
- Info
  - Project info 1
  - *New Info...*
- Results
  - *New Analysis...*
- G…s
  - **…ata 1**
  - *New Graph...*

Family

- **Data 1**
- **Data 1**

Table format: XY		X	Group A	Group B
		sensitivity	fatty_acids	Title
		X	Y	Y
1	Title	250	17.9	
2	Title	220	18.3	
3	Title	145	18.3	
4	Title	115	18.4	
5	Title	230	18.4	
6	Title	200	20.2	
7	Title	330	20.3	
8	Title	400	21.8	
9	Title	370	21.9	
10	Title	260	22.1	
11	Title	270	23.1	
12	Title	530	24.2	
13	Title	375	24.4	
14	Title			
15	Title			
16	Title			

Create New Analysis

Data to analyze

Table: Data 1

Type of analysis

Which analysis?

- **Transform, Normalize...**
  - Transform
  - Transform concentrations (X)
  - Normalize
  - Prune rows
  - Remove baseline and column math
  - Transpose X and Y
  - Fraction of Total
- **XY analyses**
  - Nonlinear regression (curve fit)
  - Linear regression
  - Fit spline/LOWESS
  - Smooth, differentiate or integrate curve
  - Area under curve
  - Deming (Model II) linear regression
  - Row means with SD or SEM
  - Correlation
  - Interpolate a sta rd curve
- **Column analyses**
- **Grouped analyses**
- **Contingency table analyses**
- **Survival analyses**

Analyze which data sets?

A:fatty_acids

When you analyze tables or graphs with more than one data set, use this space to select which data set(s) to analyze.

Select All | Deselect All

?

Cancel | OK

Parameters: Correlation

Compute correlation between which pairs of columns?

- ○ Compute r for every pair of Y data sets (Correlation matrix)
- ◉ Compute r for X vs. every Y data set:
  - X: sensitivity
- ○ Compute r between two selected data sets:
  - X: sensitivity
  - A: fatty_acids

Assume data are sampled from Gaussian distributions?

- ◉ Yes. Compute Pearson correlation coefficients
- ○ No. Compute nonparametric Spearman correlation

Options

P value: ○ One-tailed ◉ Two-tailed

Confidence interval: 95%

Output

Show this many significant digits (for everything except P values): 4

P Value Style: GP: 0.1234 (ns), 0.0332 (*), 0.0021 (**),... N= 6

Graphing

☑ Create a heatmap of the correlation matrix

☐ Make these choices the default for future analyses

? Cancel OK

correlation.pzfx — Edited

Search

- Data Tables
  - **Data 1**
  - *New Data Table...*
- Info
  - Project info 1
  - *New Info...*
- Results
  - **Correlation of Data 1**
  - *New Analysis...*
- Graphs
  - Data 1

Family

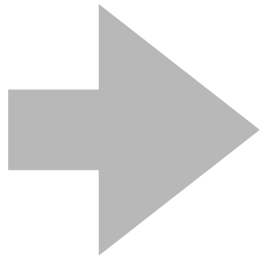
- **Data 1**
  - **Correlation**

	Correlation	A	B
		sensitivity vs. fatty_acids	Title
		Y	Y
1	**Pearson r**		
2	r	0.7700	
3	95% confidence interval	0.3804 to 0.9275	
4	R squared	0.5929	
5			
6	**P value**		
7	P (two-tailed)	0.0021	
8	P value summary	**	
9	Significant? (alpha = 0.05)	Yes	
10			
11	**Number of XY Pairs**	13	
12			
13			
14			

Spearman's rank correlation is a non-parametric measure of dependence

Spearman's ρ is just Pearson's r computed on the ranks of the x and y values which is a robust measure of correlation.

x	y
17.9	250
18.3	220
18.3	145
18.4	115
18.4	230
20.2	200
20.3	330
21.8	400
21.9	370
22.1	260
23.1	270
24.2	530
24.4	375



x rank	y rank
1.0	6.0
2.5	4.0
2.5	2.0
4.5	1.0
4.5	5.0
6.0	3.0
7.0	9.0
8.0	12.0
9.0	10.0
10.0	7.0
11.0	8.0
12.0	13.0
13.0	11.0

https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient

Parameters: Correlation

Compute correlation between which pairs of columns?

- ○ Compute r for every pair of Y data sets (Correlation matrix)
- ◉ Compute r for X vs. every Y data set:
  - X: sensitivity
- ○ Compute r between two selected data sets:
  - X: sensitivity
  - A: fatty_acids

Assume data are sampled from Gaussian distributions?

- ○ Yes. Compute Pearson correlation coefficients
- ◉ No. Compute nonparametric Spearman correlation

Options

P value: ○ One-tailed ◉ Two-tailed

Confidence interval: 95%

Output

Show this many significant digits (for everything except P values): 4

P Value Style: GP: 0.1234 (ns), 0.0332 (*), 0.0021 (**),... N= 6

Graphing

☑ Create a heatmap of the correlation matrix

☐ Make these choices the default for future analyses

? Cancel OK

Power analysis

Statistical power is the probability of detecting an effect that actually does exist.

power:
The probability of getting a statistically significant result if the null hypothesis actually is actually false.

power analysis:
The process of assigning and/or computing four quantities (sometimes more) that describe one's experiment:

1. The sample size N
2. The false positive probability α (confidence = $1 - \alpha$)
3. The false negative probability β (power = $1 - \beta$)
4. The anticipated effect size

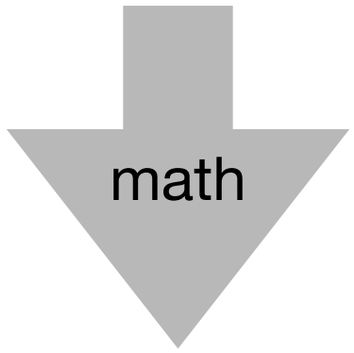
Motulsky, Ch. 20

Example: sex ratio

1. Confidence level: $1 - \alpha = 95\,\%$
2. Number of birth records: $N = 19500$
3. Hypothesized effect size: $|p(\text{boy}) - p(\text{girl})| = 2\,\%$

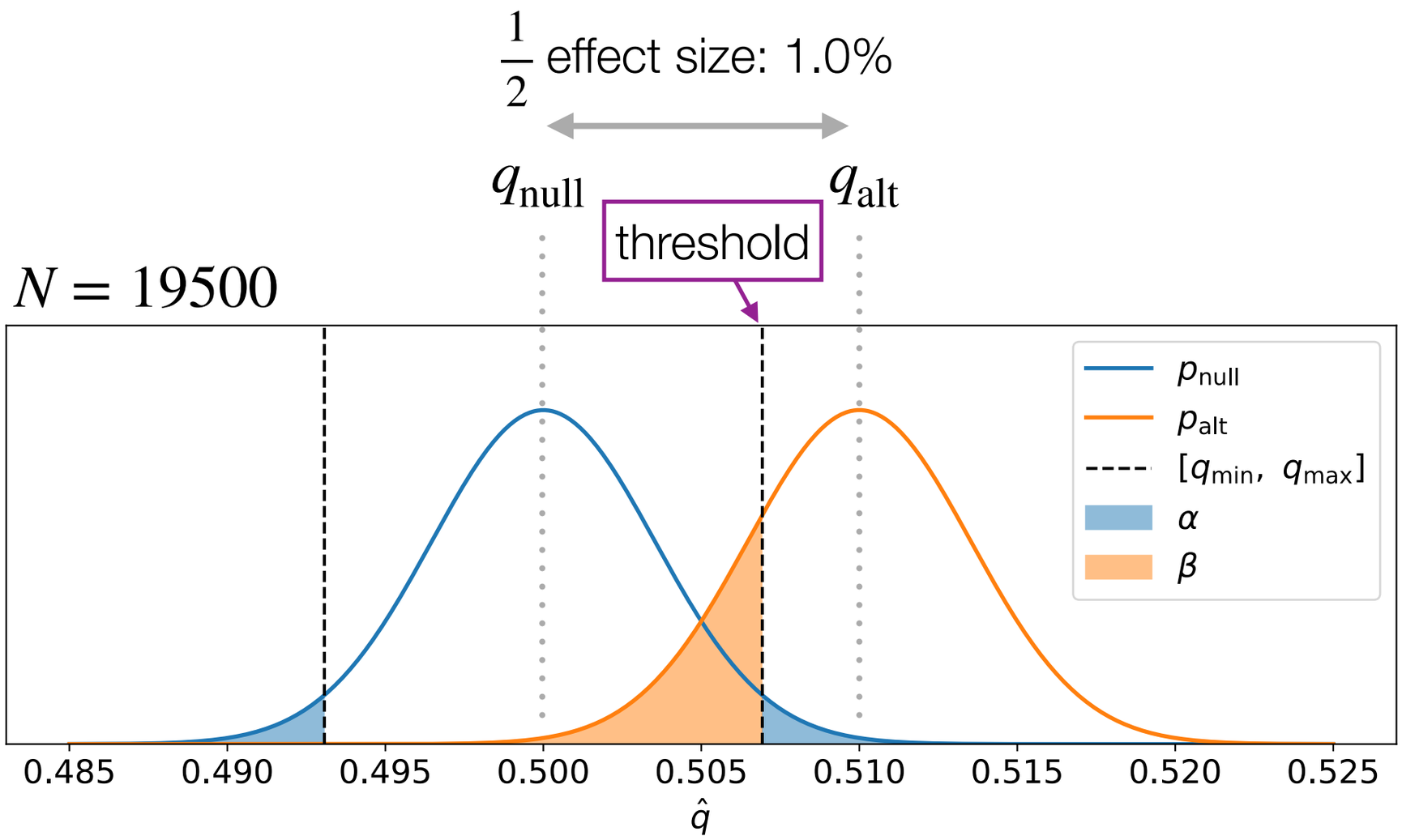
The key parameter is $q = p(\text{boy})$, so we use

$$q_{\text{null}} = 50\,\%\,, \quad q_{\text{alt}} = 51\,\%$$



4. We compute a statistical power of: $1 - \beta = 80\,\%$

Statistical power example: sex ratio data



False Positive Probability: $\alpha = 0.05$

False Negative Probability: $\beta = 0.20$
(or 80% power)

Power analysis claims come in different forms

There are four relevant parameters: N, α, β, and effect size.

Power analysis involves assuming values for any three parameters and computing the value of the forth

"Controlling the false positive rate at $\alpha = 5\,\%$, the statistical power at $1 - \beta = 80\,\%$, and assuming an effect size of $2\,\%$, our study will require using $N = 19500$ birth records."

"Using $N = 19500$ birth records, controlling the false positive rate at $\alpha = 5\,\%$, and assuming a $2\,\%$ effect size, our study will have $1 - \beta = 80\,\%$ power."

"Controlling the false positive rate at $\alpha = 5\,\%$, the statistical power at $1 - \beta = 80\,\%$, and using $N = 19500$ birth records, our study will be sensitive to an effect size of $2\,\%$."

"Using $N = 19500$ birth records, assuming an effect size of $2\,\%$, and holding the statistical power to $1 - \beta = 80\,\%$, our study will be able to hold the false positive rate to $\alpha = 5\,\%$."

What if…

What happens to the sample size if:

- SD increases
- Power increases
- Detectable difference decreases
- Level of significance decreases

You will most likely do one of these two things:

You are supposed to do this:

1. Assume a false positive rate of $\alpha = 5\,\%$ (standard)
2. Assume a power of $1 - \beta = 80\,\%$ (standard)
3. Assume what you consider to be a biologically significant effect size
4. Compute & use the required sample size N.

You'll actually probably do this:

1. Assume a false positive rate of $\alpha = 5\,\%$ (standard).
2. Assume a power of $1 - \beta = 80\,\%$ (standard)
3. Assume a reasonable / affordable sample size N
4. Compute & report the detectable effect size.

If the detectable effect size is too small

Power analysis example: body temperature

1. Assume a false positive rate of $\alpha = 5\,\%$ (standard).
2. Assume a power of $1 - \beta = 80\,\%$ (standard)
3. Assume what you consider to be a biologically significant effect size: ~~$\Delta\mu = 0.1$ F.~~ $\Delta\mu = 0.2$ F

   The key parameter is the "normalized effect size": $\dfrac{\Delta\mu}{\sigma}$

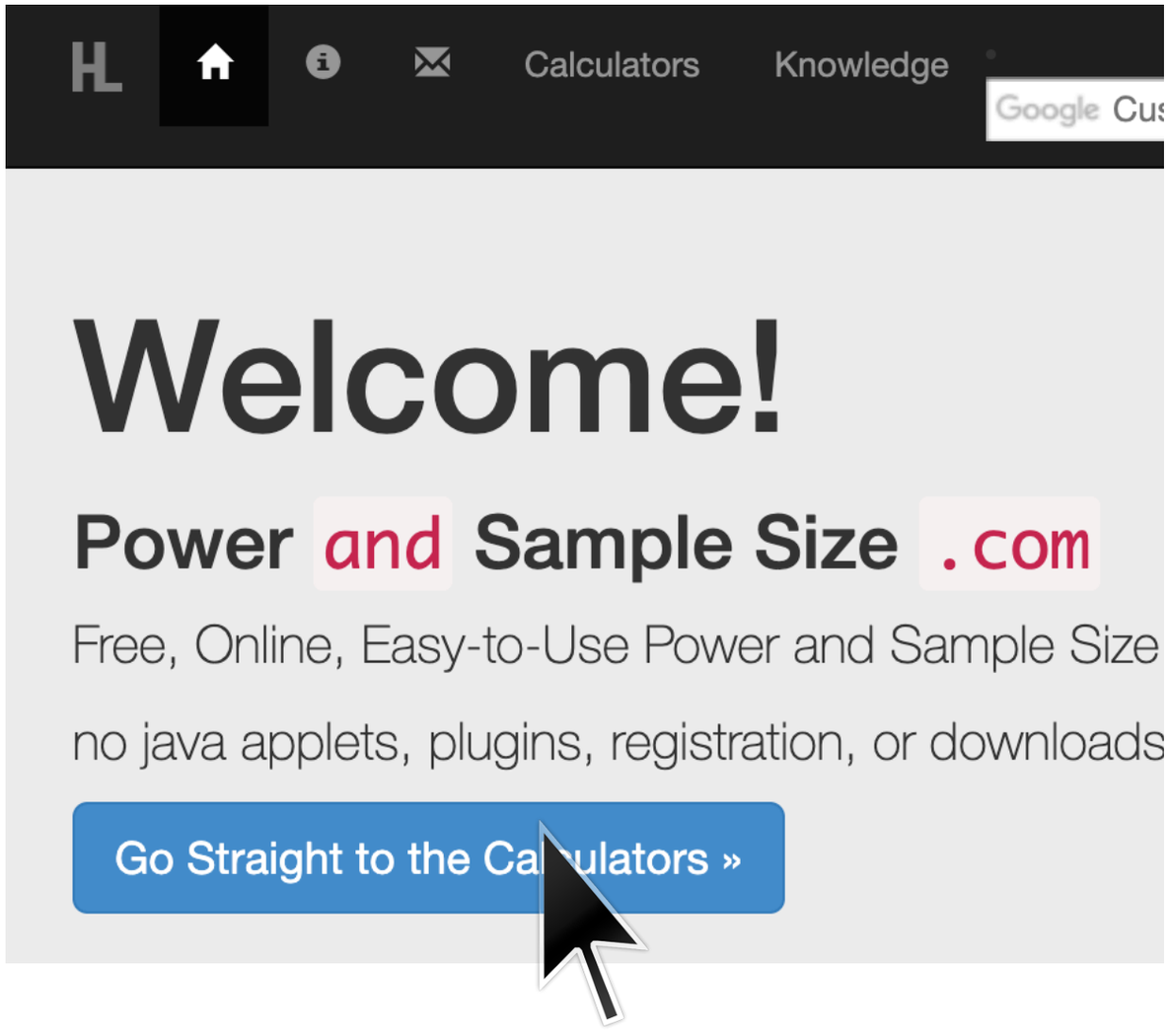
   From preliminary data, we know $\sigma \approx 0.7$ F

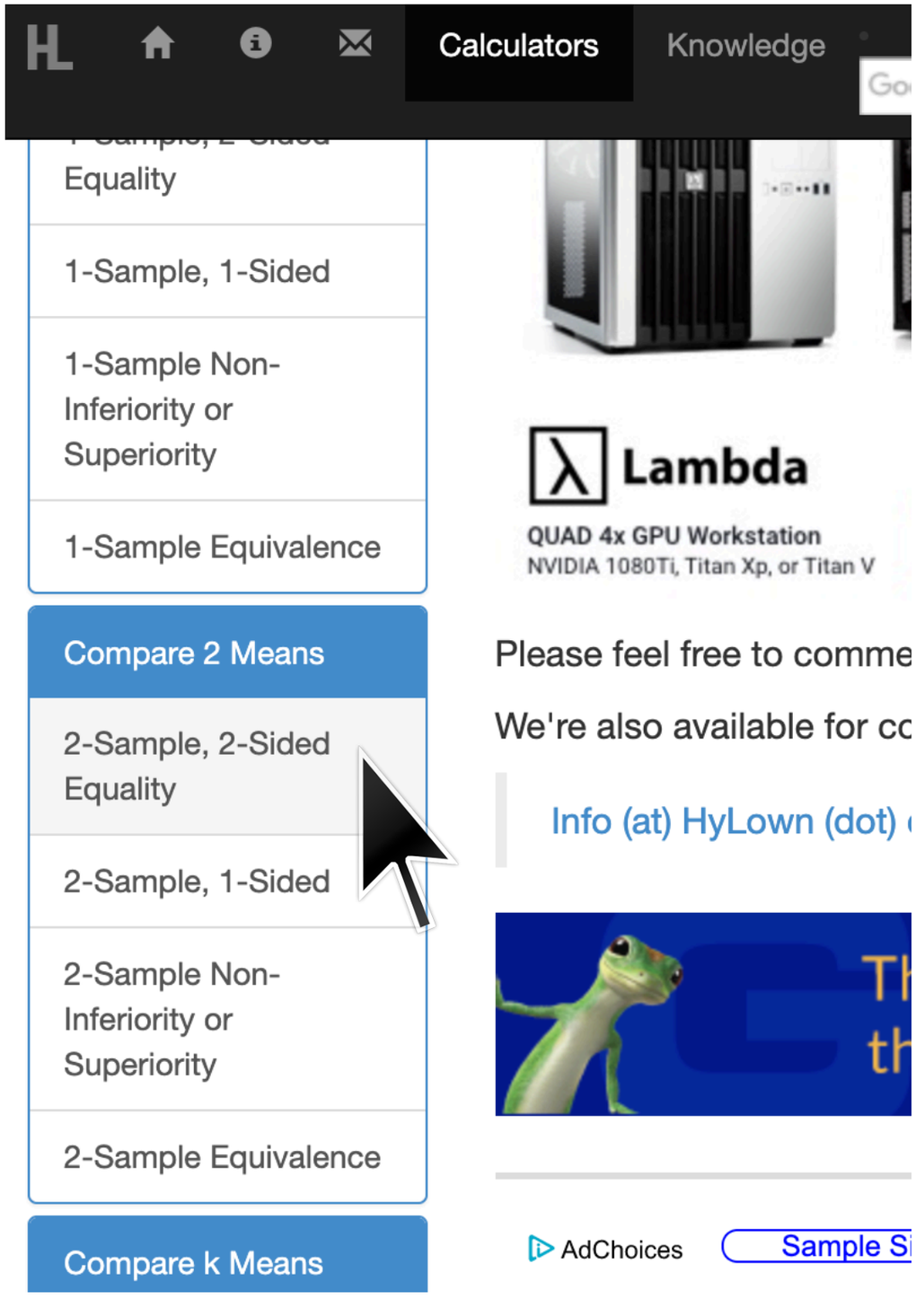
4. Compute the required sample size: ~~$N = 1540$~~ $N = 386$

   Too big! OK.

There are a number of online power analysis calculators

http://powerandsamplesize.com/





Calculate: Sample Size

Sample Size, n_B: 192

Power, $1-\beta$: 0.80

Type I error rate, α: 5%

98.1 **Group 'A' mean, μ_A**

98.3 **Group 'B' mean, μ_B**

0.7 **Standard Deviation, σ**

1 **Sampling Ratio, $\kappa = n_A/n_B$**

Calculate

Calculate: Power

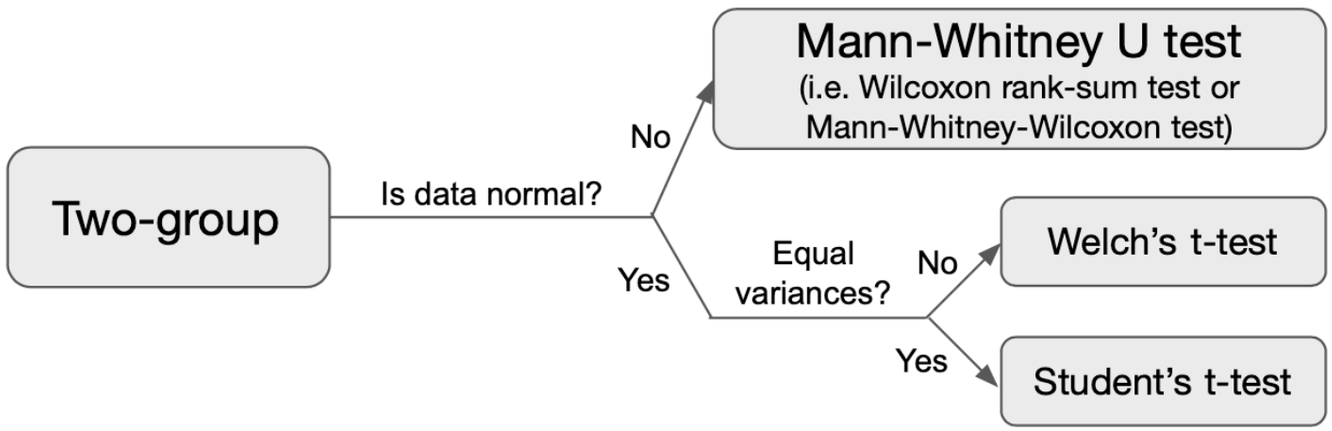
Sample Size, n_B	Power, $1-\beta$	Type I error rate, α
250	0.892	5%

98.1	**Group 'A' mean, μ_A**
98.3	**Group 'B' mean, μ_B**
0.7	**Standard Deviation, σ**
1	**Sampling Ratio, $\kappa = n_A/n_B$**

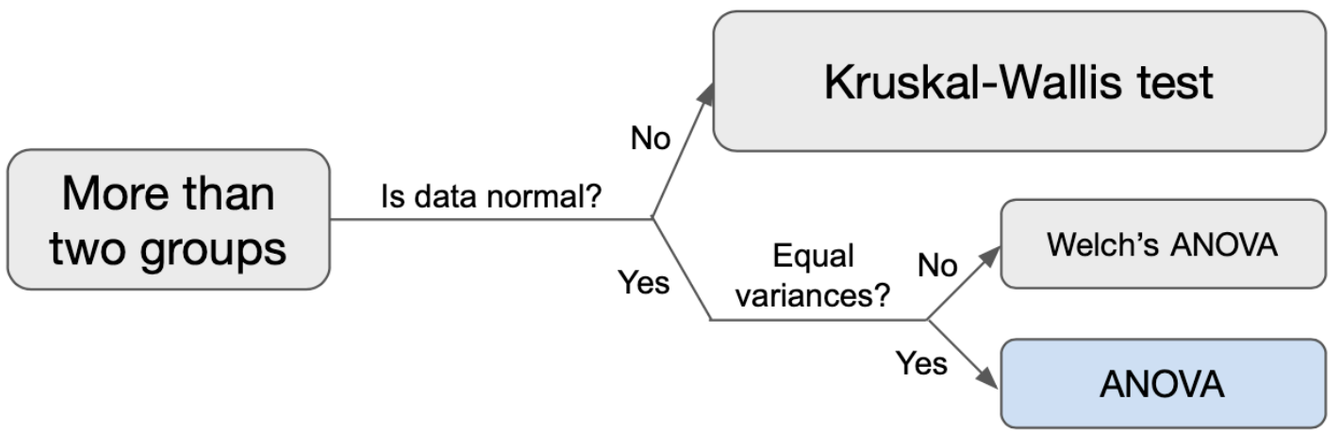
Calculate

Analysis of variance (ANOVA)

Where we stand: to compare numerical data in multiple independent groups



Assumptions:
- Errors should be random and independent
- Normality
- Homogeneity of variances



If assumptions violated,
- Transform your data and see if they meet assumptions
- If still violated, try non-parametric approach (Kruskal-Wallis test)

Fisher's solution: ANOVA (Analysis of Variance)

- **Idea:** Instead of doing multiple pairs of comparisons, why don't we do a single test?
  - This test will tell us whether there is difference in any of the means.
  - We do multiple comparisons between pairs **only after** we know there is difference in means across the groups.

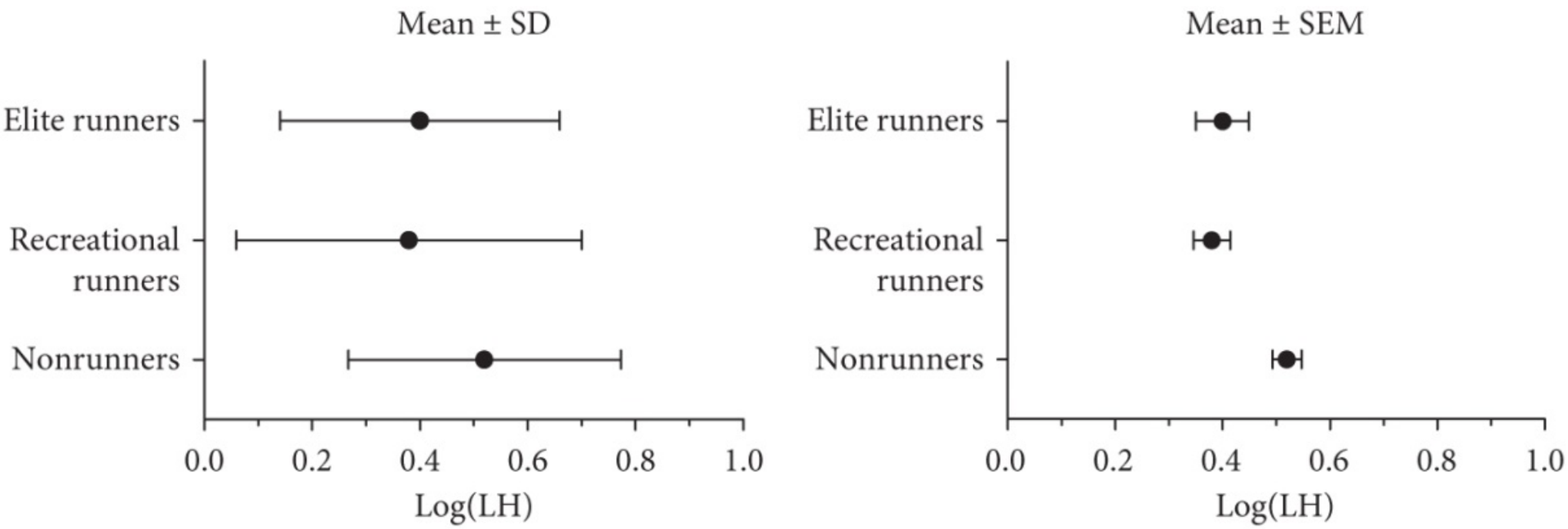
- **Hypotheses**:
  - H_0: All group means are the same. (H_0: $\mu_1 = \mu_2 = \ldots = \mu_p$)
  - H_a: At least one group mean is different.

- **Process:**
  - ($p>\alpha$) fail to reject H_0 → all group means are the same → No further investigation

  - ($p<\alpha$) reject H_0 → At least one group mean is different → Post-hoc analysis (i.e., pairwise comparison) to identify which group(s) mean(s) are significantly different.

One-way ANOVA example: hormone levels in runners

Hetland et al. (1993) investigated the level of luteinizing hormone (LH) in runners. Runners were classified into three groups: elite runners, recreational runners, and nonrunners.



GROUP	LOG(LH)	SD	SEM	N
nonrunners	0.52	0.25	0.027	88
recreational runners	0.38	0.32	0.034	89
elite runners	0.40	0.26	0.049	28

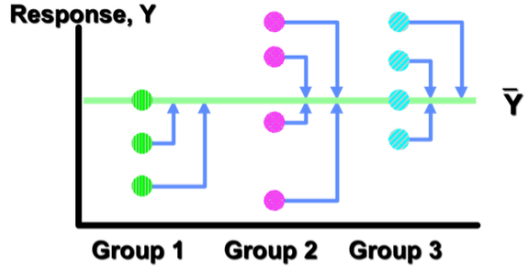
One-way ANOVA analyzes whether group means are significantly different

Null hypothesis: All group means are the same

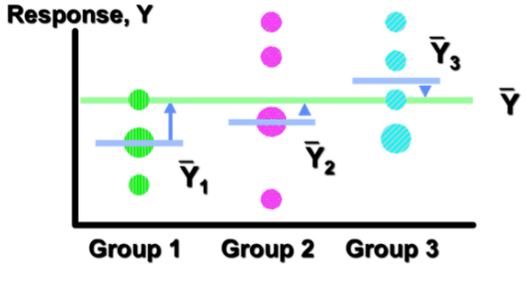
Alternative hypothesis: At least one group mean is different

SS = sum of squares

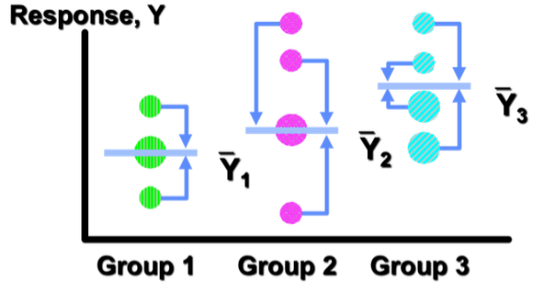
$$\underbrace{\sum_i (y_i - \hat{\mu})^2}_{\text{SS}_{\text{total}}} = \underbrace{\sum_i (y_i - \hat{\mu}_{g_i})^2}_{\text{SS}_{\text{within}}} + \underbrace{\sum_i (\hat{\mu}_{g_i} - \hat{\mu})^2}_{\text{SS}_{\text{between}}}$$



$$SS_{Total} = \Sigma_{i,j}(y_{ij} - \bar{y})^2$$



$$SS_{between} = \Sigma_i n_i(\bar{y}_i - \bar{y})^2$$



$$SS_{within} = \Sigma_{i,j}(\bar{y}_{i,j} - \bar{y_i})^2$$

One-way ANOVA analyzes whether group means are significantly different

$$\underbrace{\sum_i (y_i - \hat{\mu})^2}_{\mathrm{SS}_{\text{total}}} = \underbrace{\sum_i (y_i - \hat{\mu}_{g_i})^2}_{\mathrm{SS}_{\text{within}}} + \underbrace{\sum_i (\hat{\mu}_{g_i} - \hat{\mu})^2}_{\mathrm{SS}_{\text{between}}}$$

DF = degree of freedom

$$\mathrm{DF}_{\text{within}} = N - G, \quad \mathrm{MS}_{\text{within}} = \frac{\mathrm{SS}_{\text{within}}}{\mathrm{DF}_{\text{within}}}$$

MS = mean square

similar if null is true

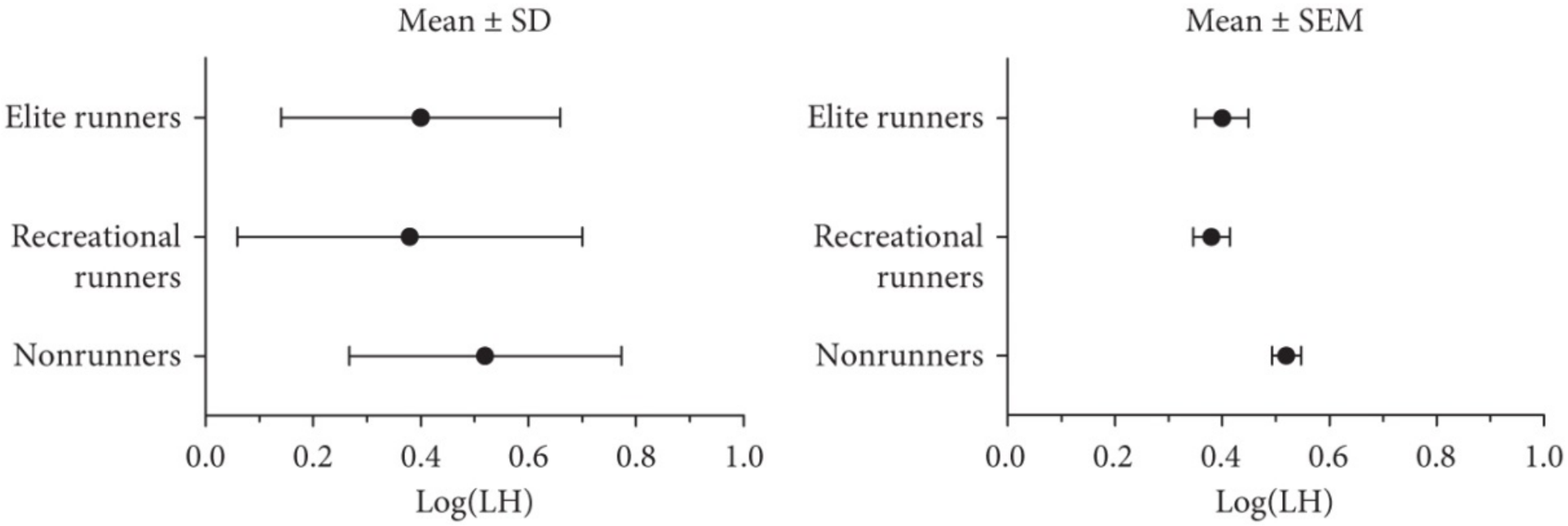
$$\mathrm{DF}_{\text{between}} = G - 1, \quad \mathrm{MS}_{\text{between}} = \frac{\mathrm{SS}_{\text{between}}}{\mathrm{DF}_{\text{between}}}$$

The corresponding F statistic is: $F = \dfrac{\mathrm{MS}_{\text{between}}}{\mathrm{MS}_{\text{within}}}$

$F \approx 1$ if null is true

The null hypothesis, implies that: $F \sim \mathrm{FDist}(\mathrm{DF}_{\text{between}}, \mathrm{DF}_{\text{within}})$

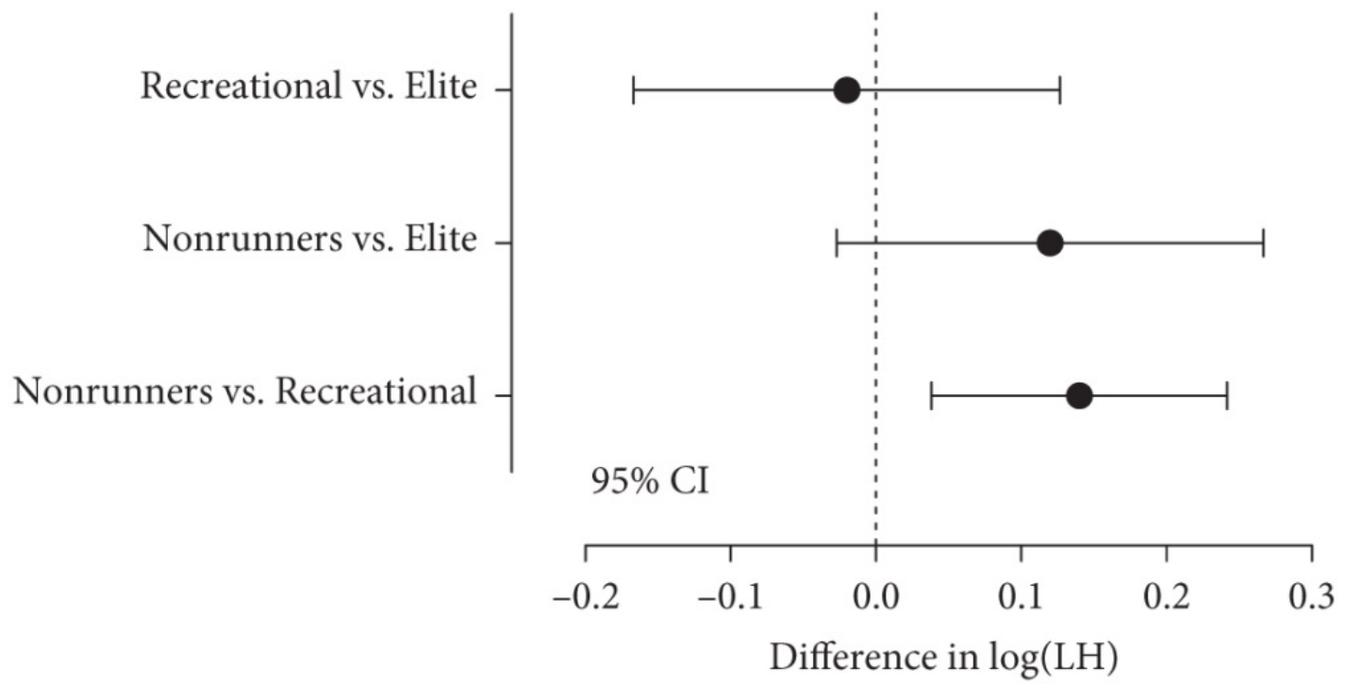
One-way ANOVA analyzes whether group means are significantly different



	SOURCE OF VARIATION	SUM OF SQUARES	DF	MS	F RATIO	P VALUE
	Between groups	0.93	2	0.46	5.69	0.0039
-	Within groups (resid.)	16.45	202	0.081		
=	Total	17.38	204			

This shows that the at least one group have significantly different mean.
It does **NOT**, however, tell which means are different. If there are differences in means, *post-hoc analysis* are typically required to identify which groups are different.

Tukey's test analyzes which pairwise comparisons in a one-way ANOVA, if any, are significant.



Tukey's test automatically incorporates the necessary multiple hypothesis correction into the test of significance.

There are other ANOVA post-hoc tests as well.

Welcome to GraphPad Prism

GraphPad Prism
Version 8.4.3 (471)

NEW TABLE & GRAPH

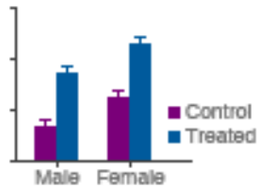
- XY
- Column
- Grouped
- Contingency
- Survival
- Parts of Whole
- Multiple variables
- Nested

EXISTING FILE

- Open a File
- LabArchives
- Clone a Graph
- Graph Portfolio

Grouped tables have two grouping variables, one defined by columns and the other defined by rows

Table format: Grouped		A			B		
		Control			Treated		
		A:Y1	A:Y2	A:Y3	B:Y1	B:Y2	B:Y3
1	Male						
2	Femal						



? Learn more

Data table:

- (•) Enter or import data into a new table
- ( ) Start with sample data to follow a tutorial

Options:

- ( ) Enter and plot a single Y value for each point
- ( ) Enter 3 replicate values in side-by-side subcolumns
- (•) Enter and plot error values already calculated elsewhere

nter: Mean, SD, N

Prism Tips

Cancel

Create

one-way_anova.pzfx

Search

- Data Tables
  - **Data 1**
  - *New Data Table...*
- Info
  - Project info 1
  - *New Info...*
- Results
  - *New Analysis...*
- Graphs
  - **Data 1**
  - *New Graph...*
- Layouts
  - *New Layout...*

Family
- **Data 1**
- **Data 1**

Table format: Grouped

		Group A			Group B			Group C			
		Nonrunners			Recreational runners			Elite runners			
		Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean
1	Title	0.52	0.25	88	0.38	0.32	89	0.4	0.26	28	
2	Title										
3	Title										
4	Title										
5	Title										
6	Title										
7	Title										
8	Title										
9	Title										
10	Title										
11	Title										
12	Title										
13	Title										
14	Title										
15	Title										
16	Title										
17	Title										
18	Title										
19	Title										
20	Title										
21	Title										
22	Title										
23	Title										
24	Title										
25	Title										
26	Title										
27	Title										

Data 1

Row 2, C: Elite runners

Create New Analysis

Data to analyze

Table: Data 1

Type of analysis

Which analysis?

- **Transform, Normalize...**
  - Transform
  - Transform concentrations (X)
  - Normalize
  - Prune rows
  - Remove baseline and column math
  - Transpose X and Y
  - Fraction of Total
- **XY analyses**
- **Column analyses**
  - t tests (and nonparametric tests)
  - One-way ANOVA (and nonparametric o...
  - One sample t and Wilcoxon test
  - Descriptive statistics
  - Normality and Lognormality Tests
  - Frequency distribution
  - ROC Curve
  - Bland-Altman method comparison
  - Identify outliers
  - Analyze a stack of P values
- **Grouped analyses**
- **Contingency table analyses**

Analyze which data sets?

- [x] A:Nonrunners
- [x] B:Recreational runners
- [x] C:Elite runners

Select All | Deselect All

?

Cancel | OK

Parameters: One-Way ANOVA (and Nonparametric or Mixed)

Experimental Design | Repeated Measures | Multiple Comparisons | Options | Residuals

Experimental design

- (•) No matching or pairing
- ( ) Each row represents matched, or repeated measures, data

	Group A	Group B	Group C	Group D
	Data Set-A	Data Set-B	Data Set-C	Title
	Y	Y	Y	Y
1				
2				
3				

Assume Gaussian distribution?

- (•) Yes. Use ANOVA.
- ( ) No. Use nonparametric test.

Assume equal SDs?

- (•) Yes. Use ordinary ANOVA test.
- ( ) No. Use Brown-Forsythe and Welch ANOVA tests.

Based on your choices (on all tabs), Prism will perform:

- Ordinary one-way ANOVA.

?

Cancel | OK

Parameters: One-Way ANOVA (and Nonparametric or Mixed)

Experimental Design | Repeated Measures | Multiple Comparisons | Options | Residuals

Followup tests

- ○ None.
- ◉ Compare the mean of each column with the mean of every other column.
- ○ Compare the mean of each column with the mean of a control column.
  - Control column: Group A: Nonrunners
- ○ Compare the means of preselected pairs of columns.
  - Selected pairs: Select...
- ○ Test for linear trend between column mean and left-to-right column order.

Which test?

Use choices on the Options tab to choose the test, and to set the defaults for future ANOVAs.

?

Cancel | OK

Parameters: One-Way ANOVA (and Nonparametric or Mixed)

Experimental Design | Repeated Measures | Multiple Comparisons | **Options** | Residuals

Multiple comparisons test

- (•) Correct for multiple comparisons using statistical hypothesis testing. Recommended.
  - Test: Tukey (recommended)
- ( ) Correct for multiple comparisons by controlling the False Discovery Rate.
  - Test: Two-stage step-up method of Benjamini, Krieger and Yekutieli (recommended)
- ( ) Don't correct for multiple comparisons. Each comparison stands alone.
  - Test: Fisher's LSD test

Multiple comparisons options

- [ ] Swap direction of comparisons (A-B) vs. (B-A).
- [x] Report multiplicity adjusted P value for each comparison.

  Each P value is adjusted to account for multiple comparisons.

Family-wise significance and confidence level: 0.05 (95% confidence interval)

Graphing

- [ ] Graph confidence intervals.
- [ ] Graph ranks (nonparametric).
- [ ] Graph differences (repeated measures).

Additional results

- [ ] Descriptive statistics for each data set.
- [ ] Report comparison of models using AICc.
- [ ] Report goodness of fit.

Output

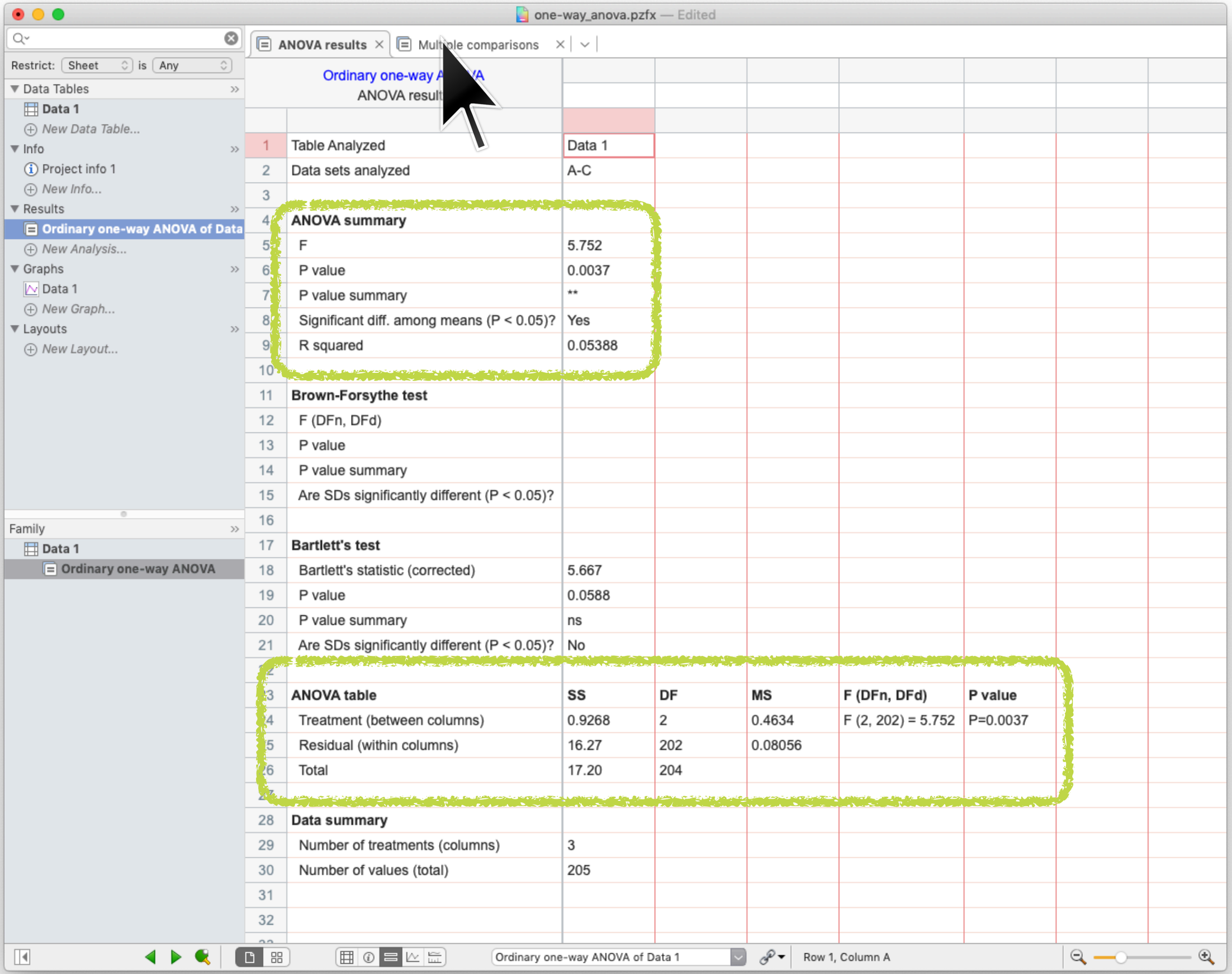
Show this many significant digits (for everything except P values): 4

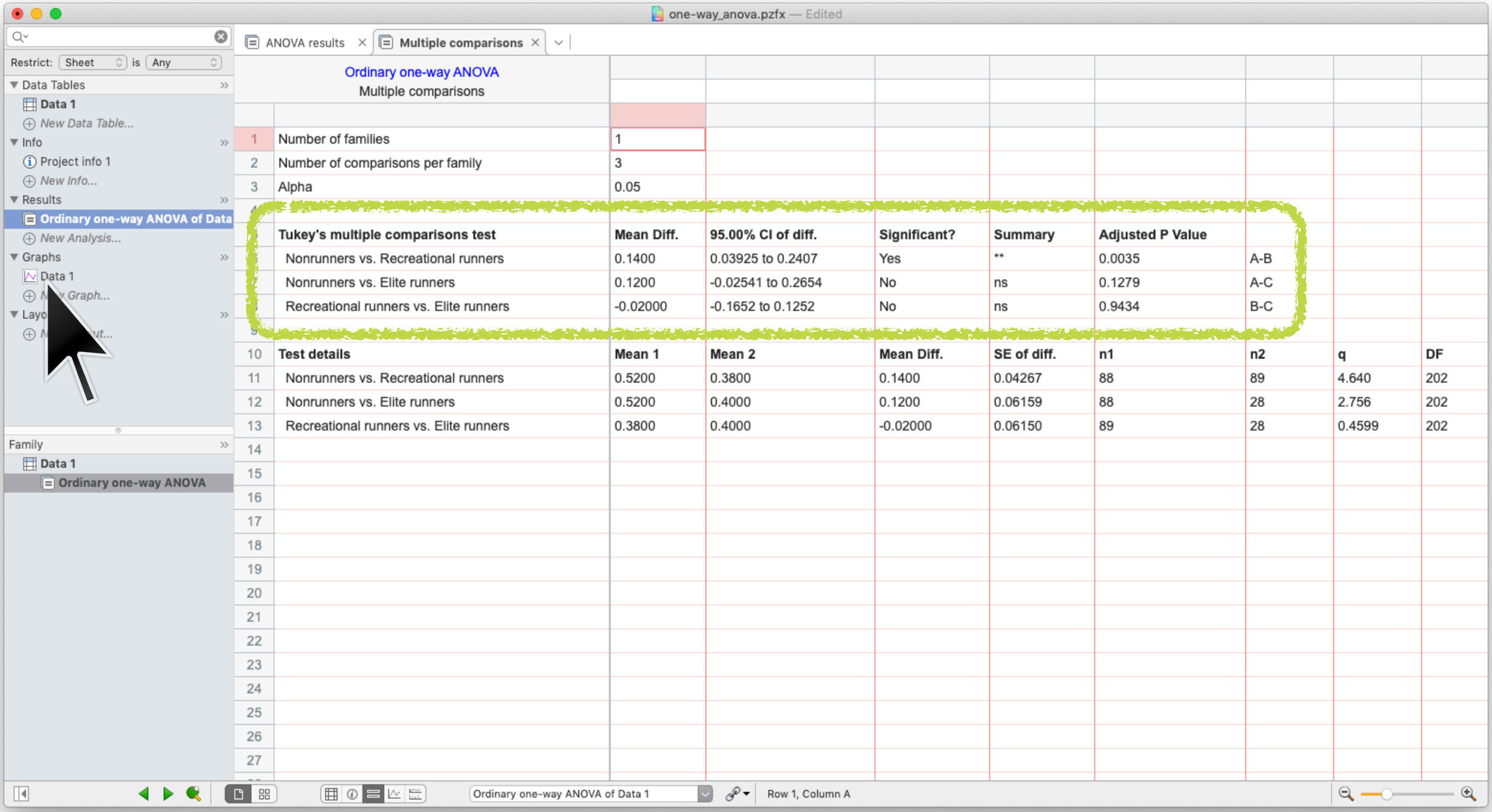
P value style: GP: 0.1234 (ns), 0.0332 (*), 0.0021 (**), 0.0002 (***), <0.0001 (**... N= 6

- [ ] Make options on this tab be the default for future One-Way ANOVAs.

?

Cancel | OK


one-way_anova.pzfx — Edited
ANOVA results
Multiple comparisons
Restrict: Sheet is Any
Data Tables
Data 1
New Data Table...
Info
Project info 1
New Info...
Results
Ordinary one-way ANOVA of Data
New Analysis...
Graphs
Data 1
New Graph...
Layouts
New Layout...
Family
Data 1
Ordinary one-way ANOVA
Ordinary one-way ANOVA
ANOVA result
1 Table Analyzed Data 1
2 Data sets analyzed A-C
3
4 ANOVA summary
5 F 5.752
6 P value 0.0037
7 P value summary **
8 Significant diff. among means (P < 0.05)? Yes
9 R squared 0.05388
10
11 Brown-Forsythe test
12 F (DFn, DFd)
13 P value
14 P value summary
15 Are SDs significantly different (P < 0.05)?
16
17 Bartlett's test
18 Bartlett's statistic (corrected) 5.667
19 P value 0.0588
20 P value summary ns
21 Are SDs significantly different (P < 0.05)? No
23 ANOVA table SS DF MS F (DFn, DFd) P value
24 Treatment (between columns) 0.9268 2 0.4634 F (2, 202) = 5.752 P=0.0037
25 Residual (within columns) 16.27 202 0.08056
26 Total 17.20 204
28 Data summary
29 Number of treatments (columns) 3
30 Number of values (total) 205
31
32
Ordinary one-way ANOVA of Data 1
Row 1, Column A


one-way_anova.pzfx — Edited
ANOVA results
Multiple comparisons
Restrict: Sheet is Any
Data Tables
Data 1
New Data Table...
Info
Project info 1
New Info...
Results
Ordinary one-way ANOVA of Data
New Analysis...
Graphs
Data 1
Family
Data 1
Ordinary one-way ANOVA
Ordinary one-way ANOVA
Multiple comparisons
Number of families 1
Number of comparisons per family 3
Alpha 0.05
Tukey's multiple comparisons test | Mean Diff. | 95.00% CI of diff. | Significant? | Summary | Adjusted P Value
Nonrunners vs. Recreational runners | 0.1400 | 0.03925 to 0.2407 | Yes | ** | 0.0035 | A-B
Nonrunners vs. Elite runners | 0.1200 | -0.02541 to 0.2654 | No | ns | 0.1279 | A-C
Recreational runners vs. Elite runners | -0.02000 | -0.1652 to 0.1252 | No | ns | 0.9434 | B-C
Test details | Mean 1 | Mean 2 | Mean Diff. | SE of diff. | n1 | n2 | q | DF
Nonrunners vs. Recreational runners | 0.5200 | 0.3800 | 0.1400 | 0.04267 | 88 | 89 | 4.640 | 202
Nonrunners vs. Elite runners | 0.5200 | 0.4000 | 0.1200 | 0.06159 | 88 | 28 | 2.756 | 202
Recreational runners vs. Elite runners | 0.3800 | 0.4000 | -0.02000 | 0.06150 | 89 | 28 | 0.4599 | 202
Ordinary one-way ANOVA of Data 1
Row 1, Column A

Change Graph Type

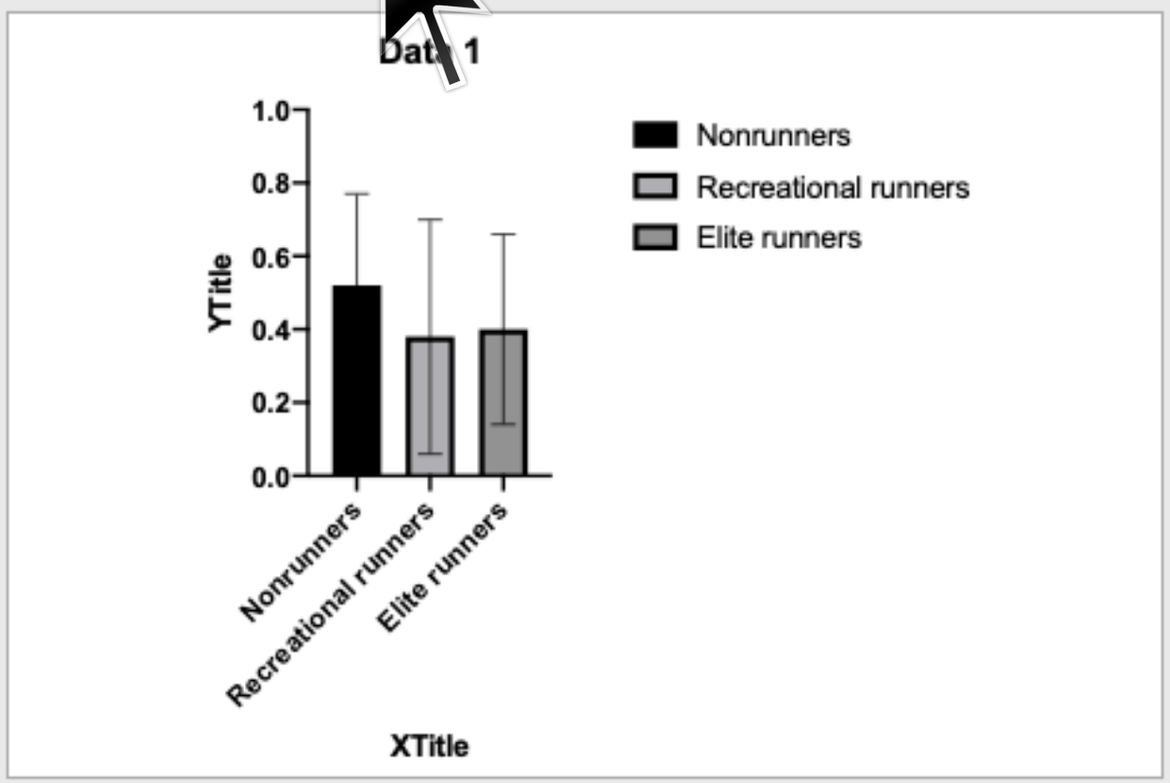
Graph family: Grouped

Individual values | Summary data | Heat Map | Three-way | Box and violin

Interleaved bars

Plot: Mean with SD

Set as default for Interleaved



?

OK

Change Graph Type

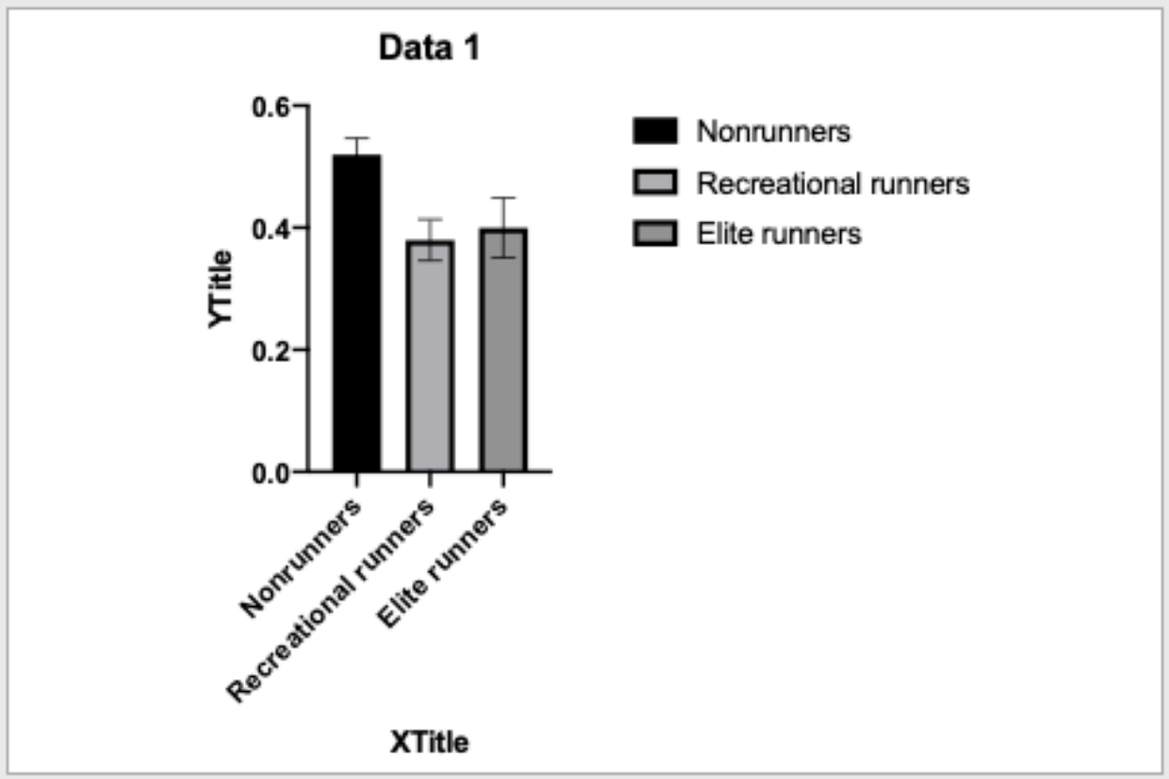
Graph family: Grouped

Individual values | Summary data | Heat Map | Three-way | Box and violin

Interleaved bars

Plot: Mean with SEM

Set as default for Interleaved bars



?

OK

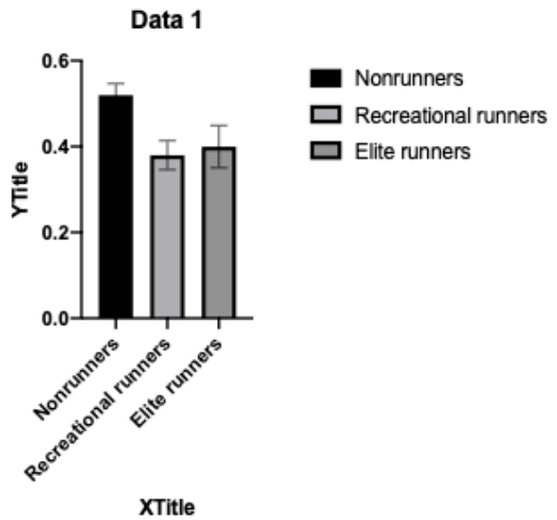
one-way_anova.pzfx — Edited

Restrict: Sheet is Any

- Data Tables
  - Data 1
  - New Data Table...
- Info
  - Project info 1
  - New Info...
- Results
  - Ordinary one-way ANOVA of Data 1
  - New Analysis...
- Graphs
  - Data 1
  - New Graph...
- Layouts
  - New Layout...

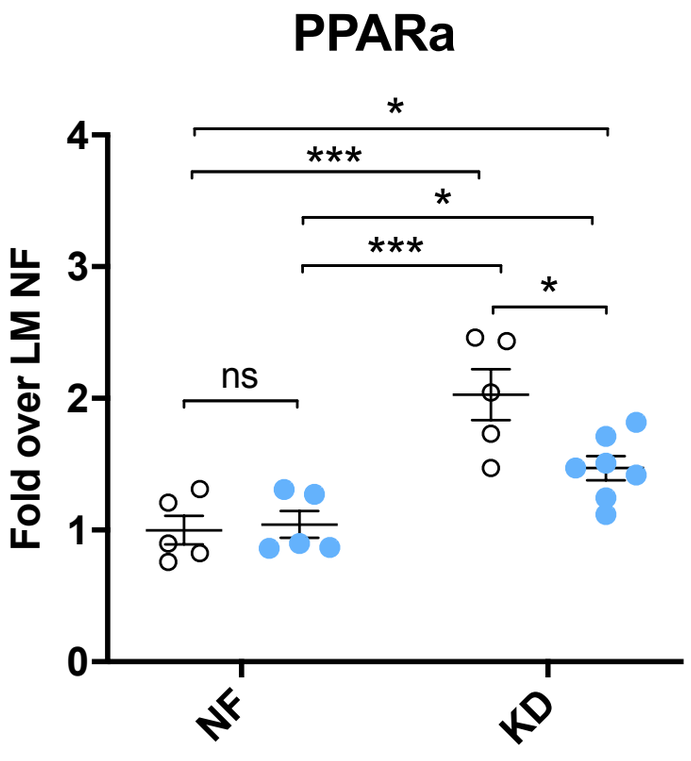
Family

- Data 1
- Data 1



Data 1

Two-way ANOVA tests whether to see if there is an interaction between groups



LM

C26

y_i = PPARa mRNA expression

x_{i1} = cancer presence (C26=tumor, LM=litter mate)

x_{i2} = food (NF=normal, KD=ketogenic)

(data curtsey of Tobias Janowitz)

Null model: $y_i = \beta_0 + \epsilon_i$

Alternative model #1: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$

Alternative model #2: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{12} x_{i1} x_{i2} + \epsilon_i$

interaction term

Welcome to GraphPad Prism

GraphPad Prism
Version 8.4.3 (471)

NEW TABLE & GRAPH

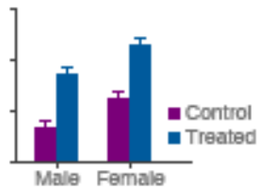
- XY
- Column
- Grouped
- Contingency
- Survival
- Parts of Whole
- Multiple variables
- Nested

EXISTING FILE

- Open a File
- LabArchives
- Clone a Graph
- Graph Portfolio

Grouped tables have two grouping variables, one defined by columns and the other defined by rows

Table format: Grouped		A			B		
		Control			Treated		
		A:Y1	A:Y2	A:Y3	B:Y1	B:Y2	B:Y3
1	Male						
2	Femal						



? Learn more

Data table:

- (•) Enter or import data into a new table
- ( ) Start with sample data to follow a tutorial

Options:

- ( ) Enter and plot a single Y value for each point
- (•) Enter 7 replicate values in side-by-side subcolumns
- ( ) Enter and plot error values already calculated elsewhere

Enter: Mean, SD, N

Prism Tips | Cancel | Create

two-way_anova.pzfx

Search

- Data Tables
  - **PPARa**
  - *New Data Table...*
- Info
  - *New Info...*
- Results
  - **2way ANOVA of PPARa**
  - *New Analysis...*
- Graphs
  - **PPARa**
  - *New Graph...*
- Layouts
  - *New Layout...*

Family

- **PPARa**
  - **2way ANOVA**
- **PPARa**

Table format: Grouped

		Group A							Group B									
		LM							C26									
		A:Y1	A:Y2	A:Y3	A:Y4	A:Y5	A:Y6	A:Y7	B:Y1	B:Y2	B:Y3	B:Y4	B:Y5	B:Y6	B:Y7	C:Y1	C:Y2	C:Y
1	NF	0.896878	0.757779	1.209183	0.824336	1.311823			0.861320	0.868462	0.899686	1.307027	1.272415					
2	KD	2.435268	2.045139	2.460515	1.472005	1.732875			1.119927	1.244879	1.509778	1.416881	1.819409	1.710366	1.470989			
3	Title																	
4	Title																	
5	Title																	
6	Title																	
7	Title																	
8	Title																	
9	Title																	
10	Title																	
11	Title																	
12	Title																	
13	Title																	
14	Title																	
15	Title																	
16	Title																	
17	Title																	
18	Title																	
19	Title																	
20	Title																	
21	Title																	
22	Title																	
23	Title																	
24	Title																	
25	Title																	
26	Title																	
27	Title																	
28	Title																	
29	Title																	
30	Title																	
31	Title																	

PPARa

Row 2, B: C26

Create New Analysis

Data to analyze

Table: PPARa

Type of analysis

Which analysis?

- **Transform, Normalize...**
  - Transform
  - Transform concentrations (X)
  - Normalize
  - Prune rows
  - Remove baseline and column math
  - Transpose X and Y
  - Fraction of Total
- **XY analyses**
- **Column analyses**
- **Grouped analyses**
  - Two-way ANOVA (or mixed model)
  - Three-way ANOVA (or mixed model)
  - Row means with SD or SEM
  - Multiple t tests - one per row
- **Contingency table analyses**
- **Survival analyses**
- **Parts of whole analyses**
- **Multiple variable analyses**
- **Nested analyses**
- **Generate curve**
- **Simulate data**

Analyze which data sets?

- A:LM
- B:C26

Select All | Deselect All

?

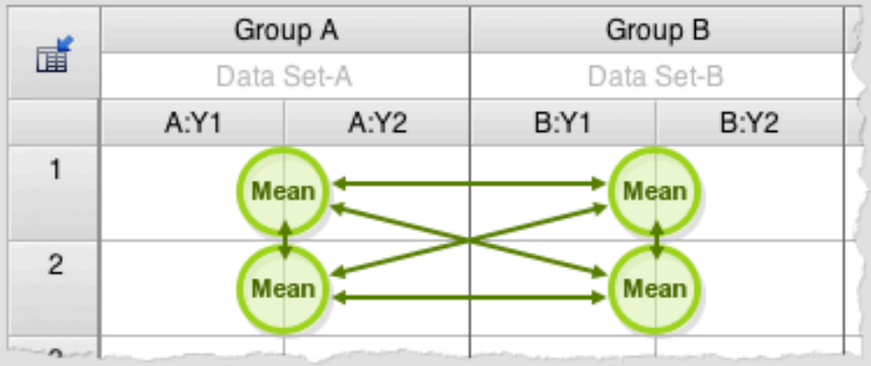
Cancel | OK

Parameters: Two-Way ANOVA (or Mixed Model)

RM Design | RM Analysis | Factor Names | **Multiple Comparisons** | Options | Residuals

What kind of comparison?

Compare cell means regardless of rows and columns



How many comparisons?

- (•) Compare each cell mean with every other cell mean.
- ( ) Compare each cell mean with the control (upper-left) cell mean.

Control cell: LM : NF

How many families?

One family for all the comparisons

Which test?

Use choices on the Options tab to choose the test, and to set the defaults for future ANOVAs.

?

Cancel | OK

Parameters: Two-Way ANOVA (or Mixed Model)

RM Design | RM Analysis | Factor Names | Multiple Comparisons | **Options** | Residuals

Multiple comparisons test

- (•) Correct for multiple comparisons using statistical hypothesis testing. Recommended.
  - Test: Holm-Sidak (more power, but can't compute confidence intervals)
- ( ) Correct for multiple comparisons by controlling the False Discovery Rate.
  - Test: Two-stage step-up method of Benjamini, Krieger and Yekutieli (recommended)
- ( ) Don't correct for multiple comparisons. Each comparison stands alone.
  - Test: Fisher's LSD test

Multiple comparisons options

- [ ] Swap direction of comparisons (A-B) vs. (B-A).
- [x] Report multiplicity adjusted P value for each comparison.

  Each P value is adjusted to account for multiple comparisons.

Family-wise significance and confidence level: 0.05

Graphing options

- [ ] Graph confidence intervals.

Additional results

- [ ] Narrative results.
- [ ] Show cell/row/column/grand predicted (LS) means.
- [ ] Report goodness of fit.

Output

Show this many significant digits (for everything except P values): 4

P value style: GP: 0.1234 (ns), 0.0332 (*), 0.0021 (**), 0.0002 (***), <0.0001 (****) N= 6

- [ ] Make options on this tab be the default for future Two-Way ANOVAs.

? Cancel OK

two-way_anova.pzfx — Edited

Search

- Data Tables
  - PPARa
  - *New Data Table...*
- Info
  - *New Info...*
- Results
  - 2way ANOVA of PPARa
  - **2way ANOVA of PPARa**
  - *New Analysis...*
- Graphs
  - PPARa
  - *New Graph...*
- Layouts
  - *New Layout...*

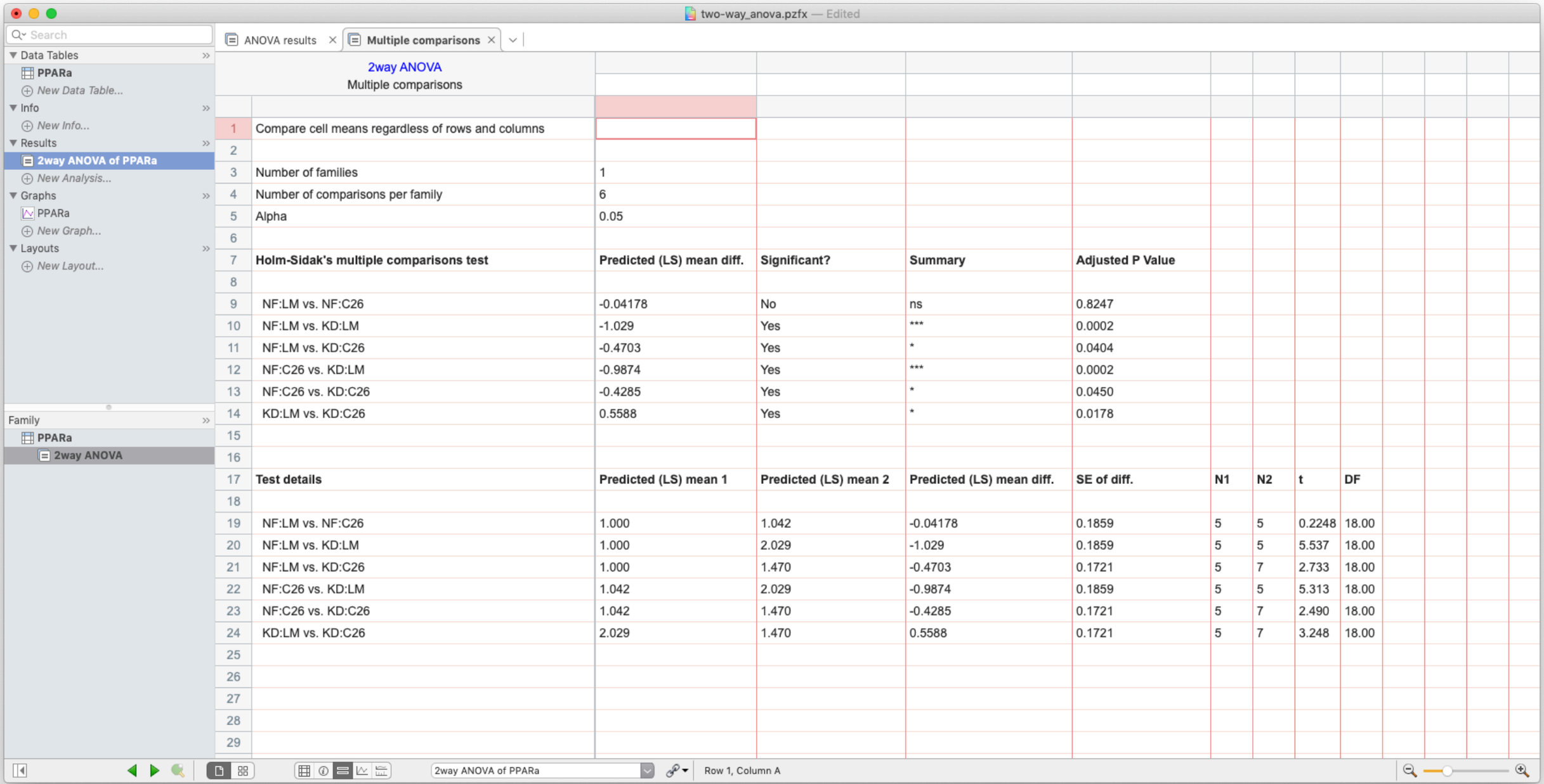
Family
- PPARa
  - 2way ANOVA

ANOVA results | Multiple comparisons

2way ANOVA
ANOVA results

1	Table Analyzed	PPARa									
2											
3	**Two-way ANOVA**	Ordinary									
4	Alpha	0.05									
5											
6	**Source of Variation**	**% of total variation**	**P value**	**P value summary**	**Significant?**						
7	Interaction	9.695	0.0291	*	Yes						
8	Row Factor	57.11	<0.0001	****	Yes						
9	Column Factor	7.185	0.0561	ns	No						
10											
11	**ANOVA table**	**SS (Type III)**	**DF**	**MS**	**F (DFn, DFd)**	**P value**					
12	Interaction	0.4856	1	0.4856	F (1, 18) = 5.623	P=0.0291					
13	Row Factor	2.860	1	2.860	F (1, 18) = 33.12	P<0.0001					
14	Column Factor	0.3599	1	0.3599	F (1, 18) = 4.167	P=0.0561					
15	Residual	1.554	18	0.08636							
16											
17	**Difference between column means**										
18	Predicted (LS) mean of LM	1.515									
19	Predicted (LS) mean of C26	1.256									
20	Difference between predicted means	0.2585									
21	SE of difference	0.1266									
22	95% CI of difference	-0.007533 to 0.5246									
23											
24	**Difference between row means**										
25	Predicted (LS) mean of NF	1.021									
26	Predicted (LS) mean of KD	1.750									
27	Difference between predicted means	-0.7288									
28	SE of difference	0.1266									
29	95% CI of difference	-0.9949 to -0.4628									

2way ANOVA of PPARa | Row 1, Column A


two-way_anova.pzfx — Edited
Search
Data Tables
PPARa
New Data Table...
Info
New Info...
Results
2way ANOVA of PPARa
New Analysis...
Graphs
PPARa
New Graph...
Layouts
New Layout...
Family
PPARa
2way ANOVA
ANOVA results
Multiple comparisons
2way ANOVA
Multiple comparisons
Compare cell means regardless of rows and columns
Number of families 1
Number of comparisons per family 6
Alpha 0.05
Holm-Sidak's multiple comparisons test | Predicted (LS) mean diff. | Significant? | Summary | Adjusted P Value
NF:LM vs. NF:C26 | -0.04178 | No | ns | 0.8247
NF:LM vs. KD:LM | -1.029 | Yes | *** | 0.0002
NF:LM vs. KD:C26 | -0.4703 | Yes | * | 0.0404
NF:C26 vs. KD:LM | -0.9874 | Yes | *** | 0.0002
NF:C26 vs. KD:C26 | -0.4285 | Yes | * | 0.0450
KD:LM vs. KD:C26 | 0.5588 | Yes | * | 0.0178
Test details | Predicted (LS) mean 1 | Predicted (LS) mean 2 | Predicted (LS) mean diff. | SE of diff. | N1 | N2 | t | DF
NF:LM vs. NF:C26 | 1.000 | 1.042 | -0.04178 | 0.1859 | 5 | 5 | 0.2248 | 18.00
NF:LM vs. KD:LM | 1.000 | 2.029 | -1.029 | 0.1859 | 5 | 5 | 5.537 | 18.00
NF:LM vs. KD:C26 | 1.000 | 1.470 | -0.4703 | 0.1721 | 5 | 7 | 2.733 | 18.00
NF:C26 vs. KD:LM | 1.042 | 2.029 | -0.9874 | 0.1859 | 5 | 5 | 5.313 | 18.00
NF:C26 vs. KD:C26 | 1.042 | 1.470 | -0.4285 | 0.1721 | 5 | 7 | 2.490 | 18.00
KD:LM vs. KD:C26 | 2.029 | 1.470 | 0.5588 | 0.1721 | 5 | 7 | 3.248 | 18.00
2way ANOVA of PPARa
Row 1, Column A

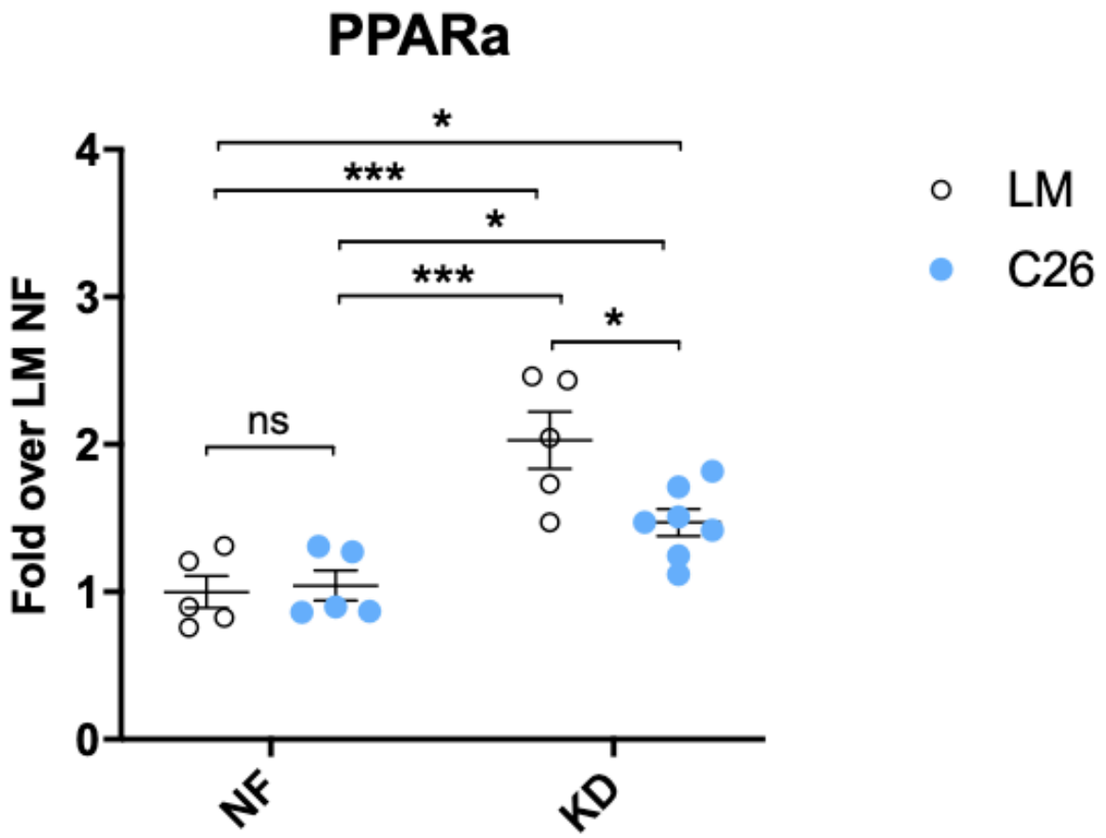
two-way_anova.pzfx — Edited

Search

- Data Tables
  - PPARa
  - *New Data Table...*
- Info
  - *New Info...*
- Results
  - 2way ANOVA of PPARa
  - *New Analysis...*
- Graphs
  - PPARa
  - *New Graph...*
- Layouts
  - *New Layout...*

Family

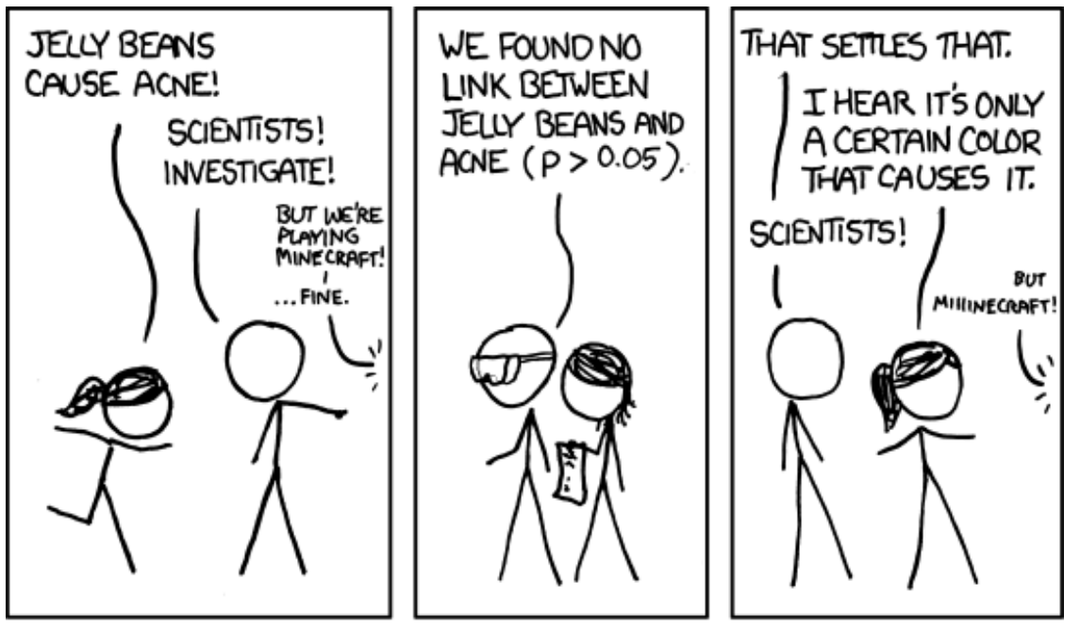
- PPARa
- PPARa

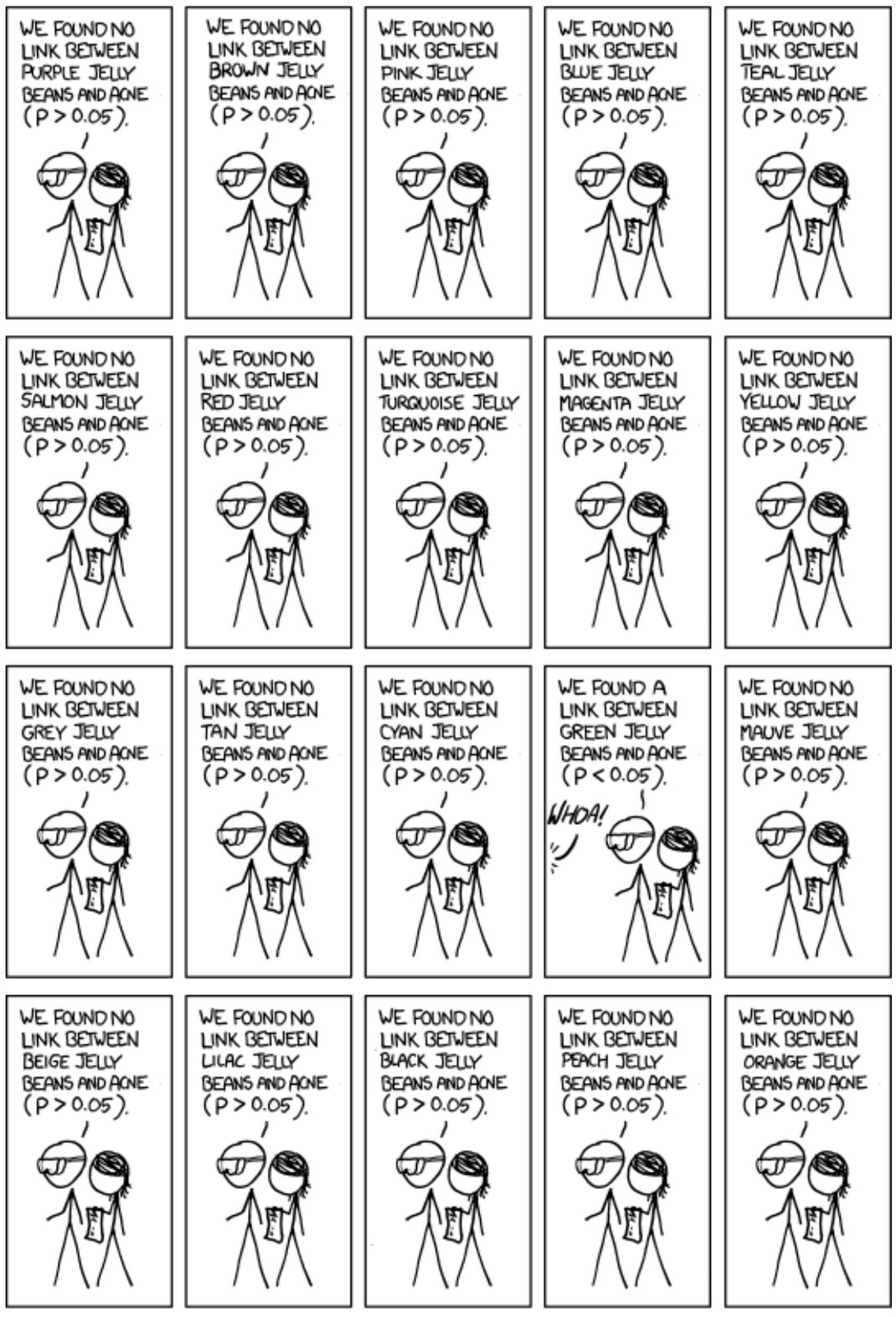


PPARa

Multiple hypothesis testing

The problem of multiple subgroups





https://xkcd.com/882/

The family-wise error rate increases rapidly with the number of tests performed

Scenario:

we perform null hypothesis tests on K independent datasets, for each of which the null hypothesis is true.

Family-wise error rate:

Probability of having at least one false positives in multiple comparisons

$$p(\text{FP} \geq 1 \,|\, \text{null hypothesis}) = 1 - \text{confidence}^K$$

FWER for different number of comparisons given different significance levels:

	1	**3**	**6**	**10**	**15**	**21**	**28**	**36**	**45**
0.05	0.05	0.14	0.26	0.4	0.54	0.66	0.76	0.84	0.90
0.01	0.01	0.03	0.06	0.1	0.14	0.19	0.25	0.30	0.36

Summary of multiple hypothesis correction techniques

Approach	What you control	Expression
No correction	α: if all null hypotheses are true, the fraction of tests that produce a significant result	$\alpha = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}$
Bonferroni / Dunn-Sidak	α: if all null hypotheses are true, the chance of obtaining one or more significant results	$\alpha = p(\#\mathrm{FP} > 0)$
False discovery rate (FDR)	Q: the fraction of all discoveries for which the null hypothesis is actually true	$Q = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TP}}$

Lucas et al. (2005)

Simple ways to counteract the multiple hypothesis problem

Bonferroni correction:

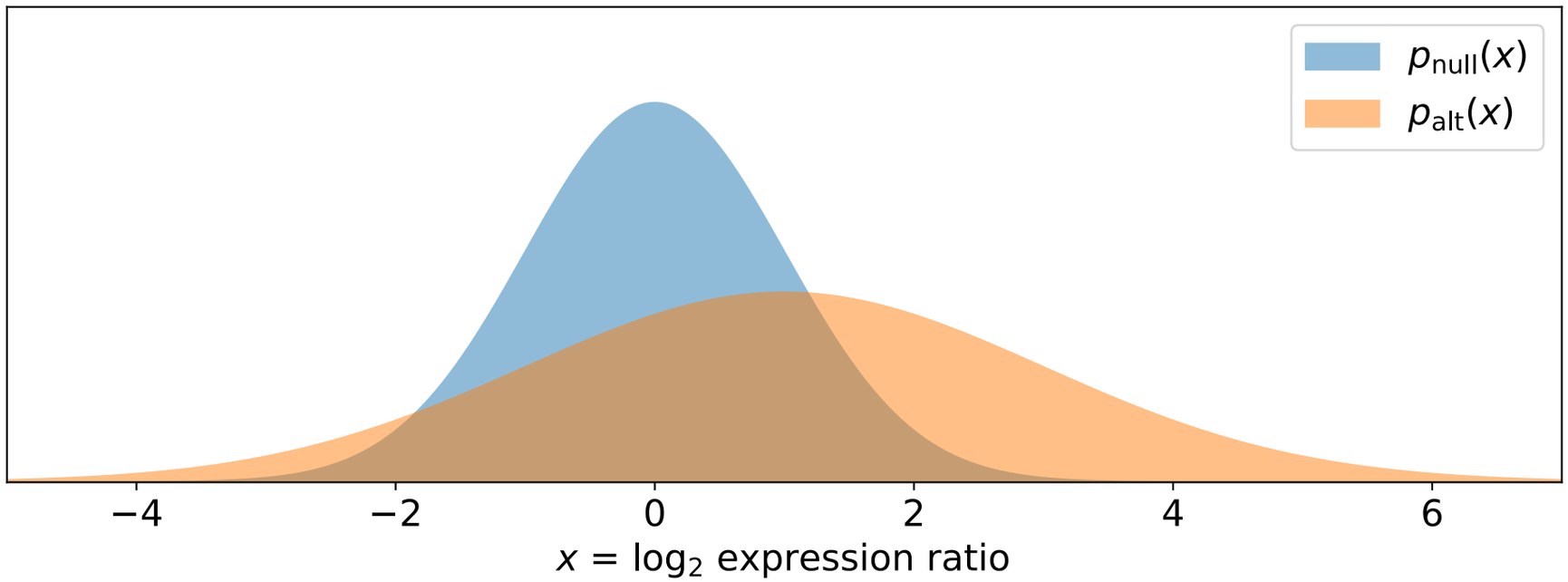
$$\alpha_{\text{Bonferroni}} = \frac{\alpha}{K}$$

Dunn-Sidak correction:

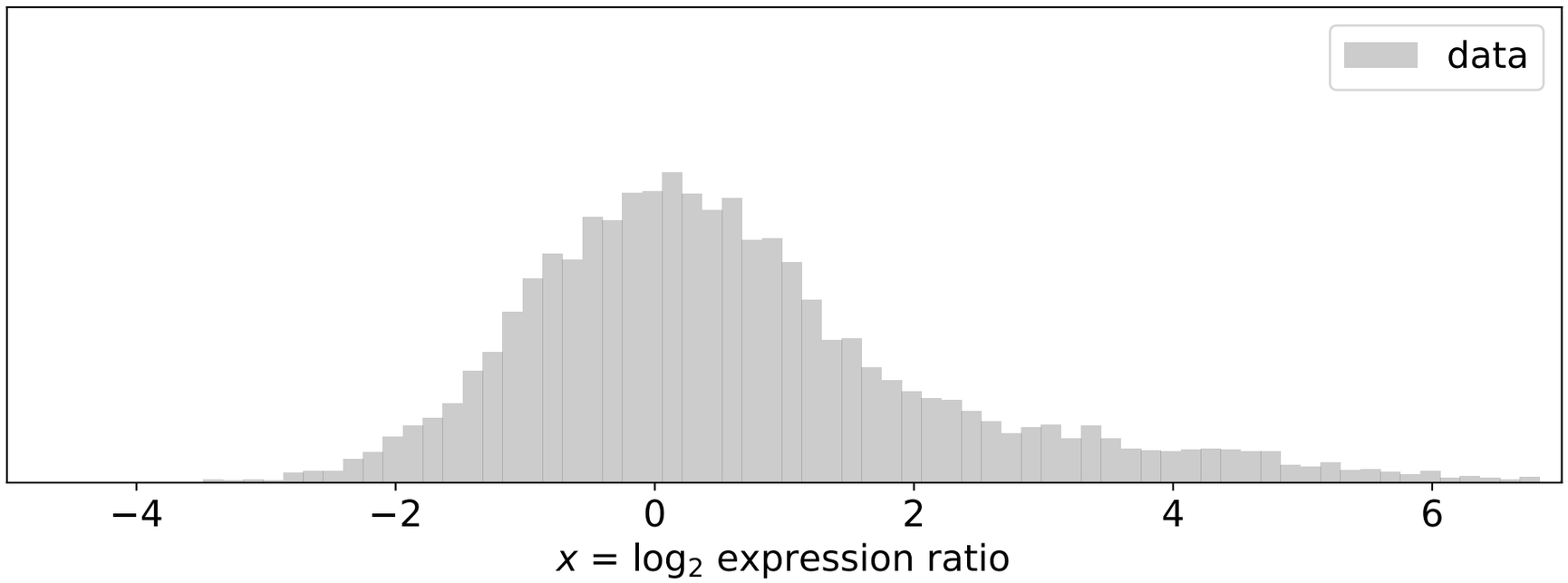
$$\alpha_{DS} = 1 - (1 - \alpha)^{1/K}$$

Dunn-Sidak is the exact solution; Bonferroni is an approximation

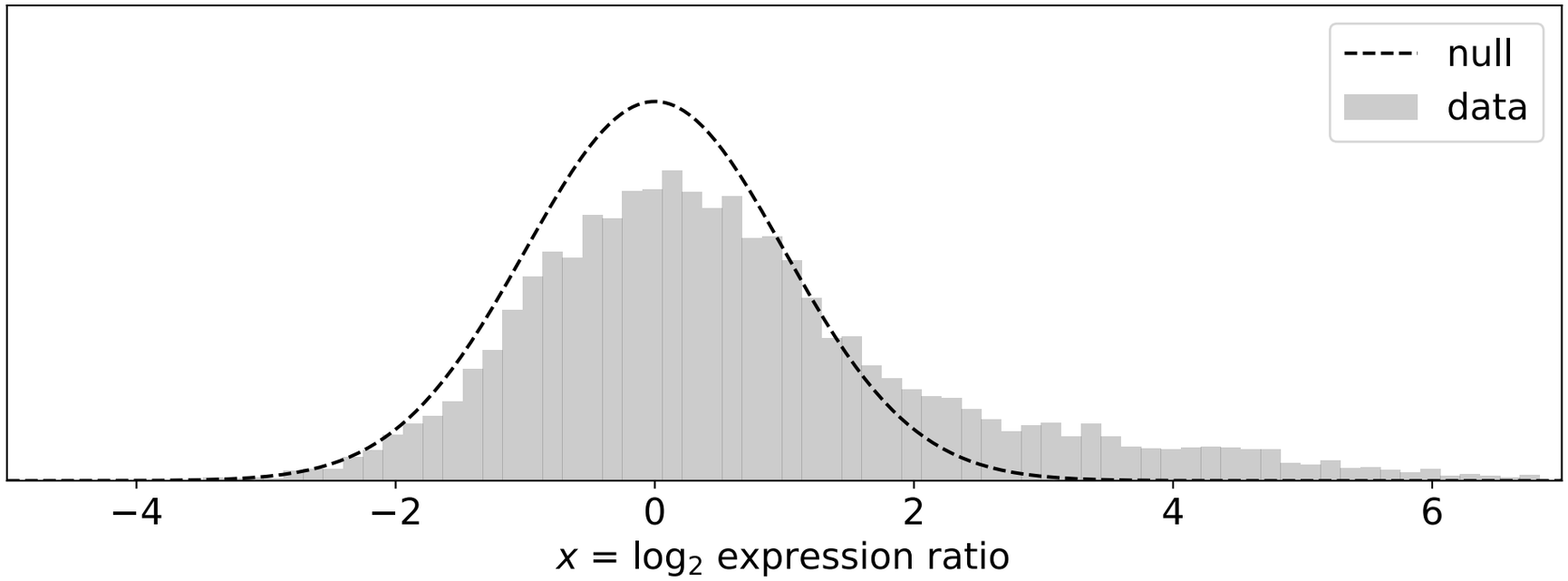
Example: differential expression (simulation)



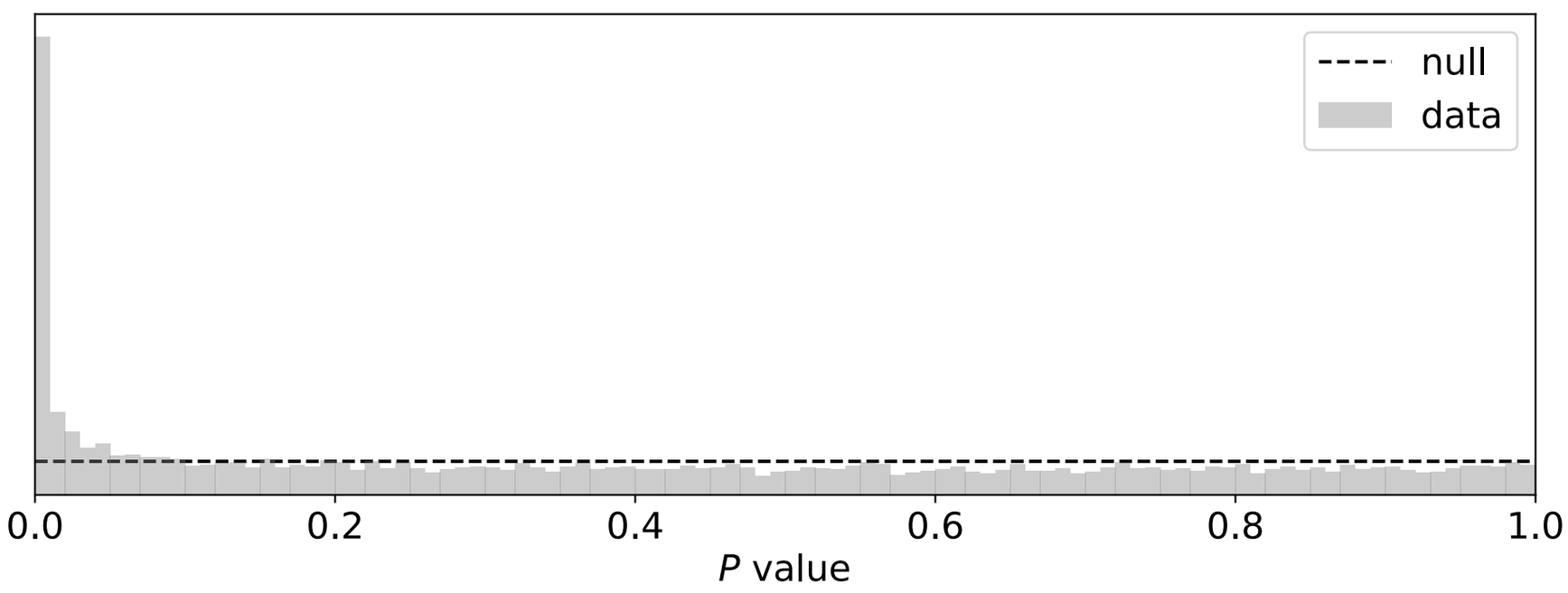
7,000 xs from $p_{\text{null}}(x)$
+ 3,000 xs from $p_{\text{alt}}(x)$



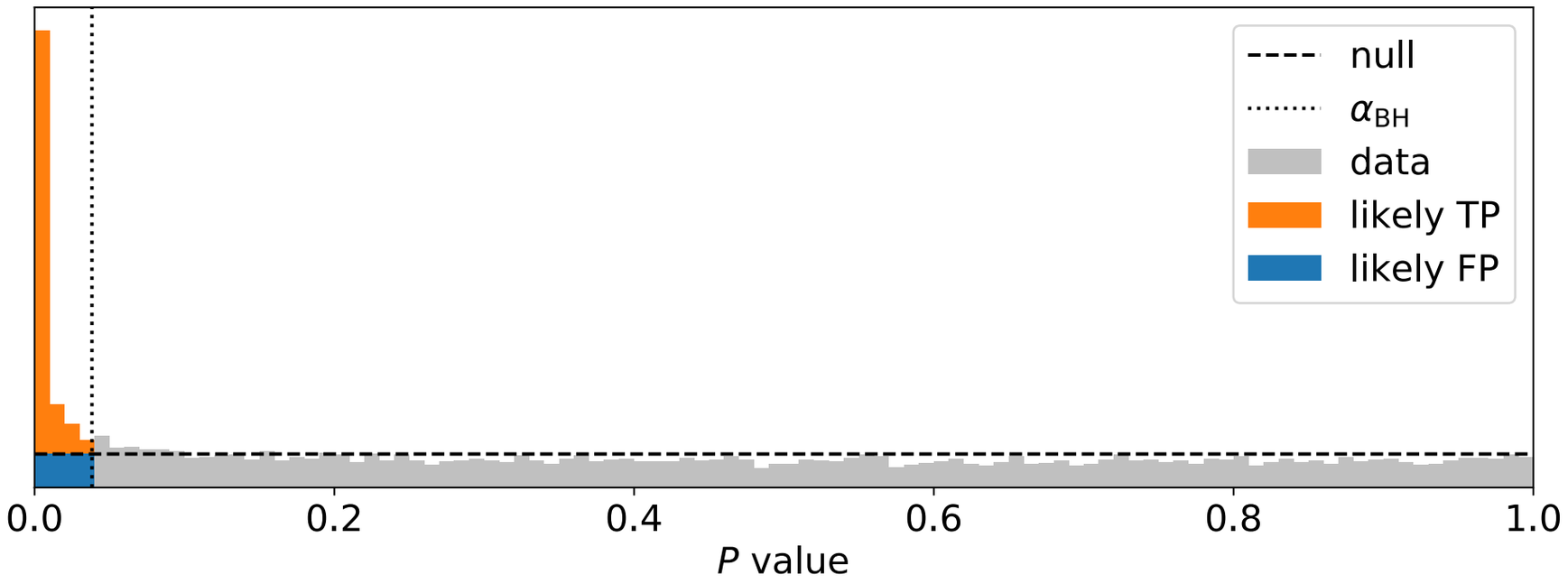
First, convert data to p-values



use knowledge of $p_{\text{null}}(x)$ to
compute a p-value for each datapoint



Benjamini–Hochberg procedure



Choose α_{BH} such to match the target False Discovery Rate (10% here):

$$\mathrm{FDR} = Q = \frac{\mathrm{FP}}{\mathrm{TP} + \mathrm{FP}} = \frac{\blacksquare}{\blacksquare + \blacksquare}$$

Declare all P-values below α_{BH} as “discoveries”.

Multiple comparisons are ubiquitous and insidious

"Most scientists are oblivious to the problems of multiplicities. Yet they are everywhere. In one or more of its forms, multiplicities are present in every statistical application. They may be out in the open or hidden. And even if they are out in the open, recognizing them is but the first step in a difficult process of inference. Problems of multiplicities are the most difficult that we statisticians face. They threaten the validity of every statistical conclusion."

from Berry (2007, p. 155), in Motulsky, Ch. 23

Multiple comparisons arise in many many contexts

multiple subgroups:
You perform tests on multiple subgroups of your data.

multiple ways to dichotomize:
You do pairwise comparisons between different combinations of subgroups.

multiple sample sizes:
You keep collecting data until you find $P < 0.05$. **DO NOT DO THIS.**

multiple ways to preprocess the data:
You analyze data preprocessed in multiple different ways.

multiple statistical tests:
You use different statistical tests on the same data before finding $P < 0.05$.

Motulsky, Ch. 23

Multiple comparisons arise in many, many contexts

multiple ways to select relevant variables:
You try to model your data using different subsets of possible variables.

multiple ways to analyze your data (“garden of forking paths”):
You try lots of qualitatively different analysis strategies.

outcome switching:
You change the quantity you care about after you’ve looked at the data.

multiple geographic areas:
E.g., you investigate a “cancer cluster” you hear about in the news.

Motulsky, Ch. 23

Correcting for multiple comparisons is not always needed

Scenario 1:
If readers can be reasonably expected to account for multiple comparisons on their own.

Scenario 2:
Before looking at the data, you have clearly defined one outcome as primary and others as secondary.

Scenario 3:
You make only a few planned comparisons and your P-values are not marginal.

Scenario 4:
A large fraction the tests you perform are significant.

Motulsky, Ch. 19

Practical advice of avoiding multiple hypothesis pitfalls

Raise your standards: use $\alpha = 0.01$, not $\alpha = 0.05$.

Separate exploratory data analysis from confirmatory data analysis.

Distinguish critical p-values from ancillary p-values.

Don't spend too much time analyzing a small dataset.

When generating small expensive datasets (e.g. mice), blind your experiments as best you can, and plan your analysis ahead of time

When in doubt, double-check your hypothesis with new data

Don't worry about informal multiple hypothesis testing when $P < 10^{-4}$.

Questions?